UNIVERSITÉ DU QUÉBEC À MONTRÉAL

BACTERIAL BIOGEOGRAPHY ACROSS BOREAL LAKES OF QUEBEC: LARGE-SCALE DRIVERS OF BACTERIAL DOMINANCE AND RARITY

DISSERTATION

PRESENTED

AS PARTIAL REQUIREMENT

OF THE DOCTORATE OF BIOLOGY

BY

JUAN PABLO NIÑO GARCIA

DECEMBER 2016

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

BIOGÉOGRAPHIE BACTÉRIENNE DES LACS BORÉAUX DU QUÉBEC: LES FACTEURS RÉGULANT LA DOMINANCE ET LA RARETÉ MICROBIENNE À GRANDE ÉCHELLE

THÈSE

PRESENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN BIOLOGIE

PAR

JUAN PABLO NIÑO GARCIA

DECEMBRE 2016

ACKNOWLEDGMENTS

I would like to start by expressing my gratitude to my advisor Paul del Giorgio for giving me the opportunity to come to Montreal and be part of this diverse and interesting research group. I have learned from you, and from the entire group, that the respect and the acceptance of the difference and the discussion are the basis for an effective collaboration in science. I feel that being part of this group has been, for me, a unique experience in science and in life. Paul, I really appreciate the amount of energy and the time that you invested on the planning of my project and on the development of my ideas during the last six years, specially because it required a lot of your patience and your understanding. I also value highly that you trusted me along the whole process, regardless the ups and downs, and that you were always confident on my personal and academic development as well as on the fulfillment of the goals that we formulated at the beginning of my Ph.D. Thanks for your orientation and guidance, I have no doubt that without them, the development of my doctoral thesis, as it is, would not be possible. Finally, I want to thank you for being always supportive and concerned when I was loosing my path, it certainly encouraged me to continue until this moment.

I also want to thank all the members of the CarBBas group that contributed to the collection and processing of the samples. Without your participation it would not be possible to reconstruct the biogeography of bacteria in boreal aquatic networks. Particularly, thanks to Nicolas Fortin-St-Gelais and Jean-François Lapierre that were always willing to listen and to discuss my ideas, especially at the early stages of my project. Nico, thanks for introducing me to the world of R and BASH coding, which open me the door to the possibility of analyzing my huge OTU tables, I really appreciate it. I want to express my gratitude to Annick St-Pierre and Alice Parkes who make our life easier by dealing with many technical, financial and logistic details, allowing us to be focused in our individual research projects. Last but not

least, thanks to Carolina García-Chaves, who collected lake Croche samples that I used for the third chapter of my thesis. Caro, we started this path together and I am happy that we managed to go trough it, helping us each other to this point when we are successfully finishing. Thanks for living this experience with me and for being ready to help me anytime.

Very special thanks go out to Clara Ruiz Gonzalez, who joined the "microbial group" about the second year of my Ph.D. and certainly became a fundamental part of the technical and scientific advance of my project, my closest scientific partner and a very good friend. Thanks for the daily discussions and conversations about our common topics of research, for helping me with the writing and the edition of all the papers that are part of this thesis, and for your pills of wisdom that helped me to overcome the challenge of learning R. I will not forget the coffee breaks, the Friday beers, las paellas, las tortillas de patatas, the polar adventures and all the fun that we had together, which made my non-scientific life in Montreal much more stimulating.

I would like also to thank David Walsh, and Steve Kembel for taking the time to participate in my evaluation committees and for helping me during my first steps of the sequencing and the analysis of the ribosomal DNA sequences. Thanks also to Carolina García-Chaves, Cynthia Soued, Rich Vogth and Nicolas Fortin-St-Gelais for helping me with the final edition of this manuscript and for critical comments and discussion.

Thanks to the Natural Science and Engineering Research Council of Canada (NSERC), Hydro-Québec, La fondation de l'UQAM and Universidad de Antioquia (Medellín-Colombia), which provided the financial support that made possible the development of my research.

A toda mi familia, gracias por su apoyo incondicional y permanente. El resultado de esta experiencia es tan suyo como mio.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GENERAL ABSTRACT

The overall objective of this thesis is to improve our understanding of the large-scale drivers and mechanisms explaining the structuring of bacterioplankton communities across boreal lakes, explicitly addressing the interplay between biogeographic processes (i.e. mass effects and species sorting) in structuring local communities. In addition, the thesis aims to discriminate between accidental and reactive taxa to further explore the drivers underlying this partition. To do so, we analyze the spatial patterns of these communities, characterized with Illumina sequencing of the 16S rRNA gen, across 386 freshwater ecosystems of boreal Quebec that cover broad ranges of environmental, hydrological and geographic conditions (Chapter I). We grouped individual OTUs into four distinct categories of spatial abundance distribution (SpAD)by fitting their abundance distributions across lakes to several statistical model distributions (normal-like, bimodal, logistic, lognormal). We then assessed the ecological features associated to these categories, including their patterns within the associated fluvial networks so as to discriminate categories that are linked to active selection within lakes, from those related to passive transport (Chapter II). Finally, we analyze the temporal dynamics of the OTUs previously grouped within the categories of SpAD and we assess the coherence between their ecological properties over space and time, in order to test whether these spatially derived properties were also maintained over time (Chapter III).

The results show that variables associated to water residence time in the landscape and water chemistry interact to shape bacterial communities along the aquatic continuum, and thus, the combined influence of hydrology, network position, and local environmental conditions structure local aquatic communities from highly diverse assemblages in headwater streams, to larger rivers and lakes dominated by fewer taxa, regardless of geographic location (Chapter I). We also show that environmental sorting and mass effects both occur across the whole aquatic network but that there is a gradient of decreasing mass effects and increasing environmental sorting towards downstream ecosystems. Thus, hydrology and network position modulate the relative influence of these biogeographic processes on taxonomic composition at any given location, likely by defining both the time frame for bacterial growth and the composition of the immigrant pool along the aquatic continuum.

The interaction between local sorting and mass effects processes was also reflected on the structure of lake bacterial communities, which were mostly composed by a small core of reactive taxa, linked to in-lake selection and recruitment, and a very large fraction of random bacteria, mostly related to passive transport (Chapter II). Interestingly, these components of lake bacterial communities were discriminated based on a small set of discrete SpAD categories that showed dramatic differences in their ranges of abundance, occurrence, and environmental tolerance in lakes as well as differences

in their local dynamics within associated riverine networks, regardless of taxonomic identity (Chapter II). Whereas normal-like and bimodal categories included the most abundant and ubiquitous taxa with clear reactive responses to environmental gradients, logistic and lognormal categories grouped mostly rare bacteria whose spatial patterns reflect hydrologic passive transport driven by the flow of water along the aquatic continuum.

We found a remarkable coherence between the SPAds and the temporal dynamics of the categories mentioned above. In this regard, we observed that most of the OTUs within the random categories of SPADs (i.e. logistic and lognormal) were not responsive to temporal changes in lake conditions over time (Chapter III), suggesting that the vast majority of rare bacterial taxa in lakes is random and driven by passive transport. Although, we also found a small group of ubiquitous and responsive rare taxa (Chapter II) and a few temporal shifters, that were spatially rare but which became abundant over time in lake Croche (Chapter III), our results suggest that the functional core of these lake communities is dominated by an adaptive group of normal-like and bimodal OTUs that respond to spatial and temporal environmental gradients through strong shifts in their relative abundances.

Overall, the results presented on this thesis demonstrate that the structure of a given bacterial community along the hydrologic continuum results from the interaction between mass effects and local environmental sorting. However, the outcome of the structuring process varies depending on the position of the system in the network, which in turns defines the rate of immigration from upstream ecosystems and the intensity of the local environmental selection. Thus, the complexity of the interactions between the described biogeographic processes together with the intrinsic eco-physiological traits of individual taxa within a community, determine the resulting large-spatial patterns across the freshwater network. The approach followed on this thesis allowed us to assess the interplay between different large-scale biogeographic processes on the distribution of environmentally reactive versus non-reactive, accidental taxa, based on their spatial behaviors and thus, it contributes not only to our understanding of the factors that shape aquatic bacterial communities, but also to those that describe the origin, structure, and ecological role of the vast rare biosphere that populates all aquatic bacterial assemblages.

Key words: aquatic bacterial communities; biogeography; boreal lakes; freshwater ecosystems; hydrologic network; abundance distributions; recruitment; species sorting; mass effects; temporal-spatial relationship; rare biosphere.

RÉSUMÉ GÉNÉRAL

L'objectif général de cette thèse est d'améliorer la compréhension des facteurs et mécanismes qui régulent la structure des communautés de bactérioplancton à grande échelle dans les lacs boréaux, en abordant explicitement l'interaction entre les processus biogéographiques (l'effet de masse et la sélection environnementale des espèces) et la structure des communautés locales. De plus, cette thèse vise à différencier les taxa accidentels des taxa réactifs pour explorer les facteurs qui sous-tendent de cette répartition. Pour cela, nous avons analysé les patrons spatiaux de ces communautés caractérisées par séquençage Illumina du gène ARNr 16S, à travers 386 écosystèmes d'eau douce boréales du Québec qui couvrent une large gamme de conditions environnementales, hydrologiques, et géographiques (Chapitre 1). Nous avons réparti les UTO dans quatre catégories de répartition spatiale de l›abondance (SpAD) en comparant la distribution de leurs abondances à différentes distributions statistiques (normale, bimodale, logistique, log normale). Nous avons ensuite évalué les traits écologiques associés à ces catégories, incluant leurs comportements au sein du réseau fluvial pour différencier entre les catégories associées à la sélection active dans les lacs et celles liées au transport passif (Chapitre 2). Enfin, nous avons analysé la dynamique temporelle des UTO précédemment regroupés dans les catégories de SpAD, et nous avons évalué la cohérence entre leurs propriétés écologiques à travers le temps et l'espace dans le but de tester si les propriétés dérivées de l'analyse spatiale sont maintenues à travers le temps.

Les résultats montrent que les variables associées au temps de résidence de l'eau dans le paysage interagissent avec la chimie de l'eau pour réguler les communautés bactériennes le long du continuum aquatique, et donc, l'influence combinée de l'hydrologie, la position dans le réseau fluvial et les conditions environnementales locales structurent les communautés aquatiques locales, allant d'assemblages très diversifiés dans les cours d'eau en amont jusqu'aux communautés dominées par peu de taxa dans les grandes rivières et les lacs en aval, indépendamment de leur position géographique (Chapitre 1). Nous avons aussi montré que l'effet de masse et la sélection environnementale des espèces surviennent le long du réseau fluvial, mais ils suivent un gradient selon lequel l'effet de masse diminue au profit de la sélection environnementale des espèces d'amont en aval. Ainsi, l'hydrologie et la position dans le réseau modulent l'influence relative de ces processus biogéographiques sur la composition taxonomique à un endroit donné, probablement en définissant le temps de croissance bactérienne ainsi que la composition du groupe d'immigrants le long du continuum aquatique.

L'interaction entre la sélection locale et l'effet de masse se reflète aussi dans la structure des communautés bactériennes lacustres, qui étaient principalement composées d'un petit noyau de taxa réactifs, lié à la sélection active et au recrutement dans les lacs, et d'une grande fraction de taxa accidentels, essentiellement liés au transport pas-

sif. (Chapitre 2). De plus, les composantes des communautés bactériennes lacustres se distinguent basées sur un nombre limité de catégories de SpAD qui manifestent de grandes différences au niveau de leur gamme d'abondance, d'occurrence, de leur tolérance environnementale dans les lacs, ainsi que des différences dans leurs dynamiques locales au sein des réseaux fluviaux, indépendamment de leur identité taxonomique (Chapitre 2). Tandis que les catégories de distributions normales et bimodales incluent les taxa les plus communes et abondantes qui présentent des réponses réactives claires face aux gradients environnementaux. Les catégories logistiques et log normale regroupent principalement les bactéries rares dont les patrons spatiaux reflètent le transport passif hydrologique dû à l'écoulement de l'eau le long du continuum aquatique.

Nous avons constaté une cohérence remarquable entre le comportement spatiale et la dynamique temporelle des catégories mentionnées ci-dessus. En ce sens, nous avons observé que la majorité des UTO dans les catégories accidentelles de comportement spatial (logistique et log normale) ne répondaient pas aux changements temporels des conditions environnementales dans les lacs (Chapitre 3), ce qui suggère que la majorité des taxa de bactéries rares dans les lacs est accidentelle et due au transport passif. Cependant, nous avons aussi trouvé un petit groupe de taxa rares omniprésents et réactifs (Chapitre 2), et quelques taxa spatialement rares mais qui sont devenus abondants au cours du temps dans le lac Croche (Chapitre 3). Nos résultats suggèrent que le noyau fonctionnel réactif de ces communautés lacustres est dominé par un groupe adaptatif d'UTO normale et bimodale qui répondent aux gradients environnementaux spatiaux et temporels à travers d'importants changements dans leurs abondances relatives.

En conclusion, les résultats présentés dans cette thèse démontrent que la structure d'une communauté bactérienne le long du continuum hydrologique résulte de l'interaction entre l'effet de masse et la sélection environnementale locale. Cependant, l'issue de ce processus de structuration varie dépendamment de sa position dans le réseau, ce dernier définissant le taux d'immigration provenant des systèmes en amont et l'intensité de la sélection environnementale locale. Par conséquent, la complexité des interactions entre les processus géographiques et les traits éco-physiologiques de chaque taxon détermine les patrons spatiaux à grande échelle à travers le réseau d'eaux douces. L'approche adoptée dans cette thèse nous a permis d'évaluer l'interaction entre différents processus biogéographiques sur la distribution de taxa réactifs versus non-réactifs, et accidentels, en se basant sur leur comportement spatial. Donc, cela contribue non seulement à la compréhension des facteurs qui façonnent les communautés bactériennes aquatiques, mais aussi à la description de l'origine, la structure et le rôle écologique de la vaste biosphère de taxa rare qui peuplent tous les assemblages bactériens aquatiques.

Mots - clés: communautés de bactérioplancton; processus biogéographiques; lacs boréaux; l'effet de masse; sélection environnementale des espèces.

INTRODUCTION

0.1 The biogeography of bacteria: drivers and mechanisms

Bacteria are highly abundant everywhere in nature, inhabiting all kinds of environments, ranging from the vast oceans to freshwaters, soils or the unusual hot springs (Torsvik, Øvreås, & Thingstad, 2002; van der Heijden, Bardgett, & van Straalen, 2008). They comprise a global combined biomass that potentially influences the balance of major elements (Whitman, Coleman, & Wiebe, 1998), and owing to their unparalleled genetic, metabolic and taxonomic diversity, they are considered as major players in biogeochemical processes that sustain life on Earth (Falkowski, Fenchel, & Delong, 2008; Newman & Banfield, 2002). In particular, the activity of bacteria is a key factor regulating ecosystem carbon dynamics, through their consumption of organic carbon, synthesis of biomass and and respiration (Cotner & Biddanda, 2002; del Giorgio & le B Williams, 2005; Pernthaler & Amann, 2005). For example, bacterioplankton respiration of terrestrial dissolved OC (DOC) has recently been shown to be one of the major pathways fuelling CO_2 emissions from inland waters (McCallister & Giorgio, 2008). These ecosystem processes are always measured as aggregated properties of bacterial communities (i.e., considering the bacterial communities as a whole), yet their wide ranges of variation (del Giorgio & Cole, 1998; Gasol & del Giorgio, 1997; Pernthaler & Amann, 2005) are likely related, at least in part, to the physiologic and taxonomic structure of the communities involved. Not all bacteria within communities are equally active in carbon consumption, respiration, or biomass production (Alonso-Sáez & Gasol, 2007; Cottrell & Kirchman, 2000; Salcher, 2013) and a large fraction of bacteria can be dormant or even dead (Carini et al., 2016; Jones & Lennon, 2010; Lennon & Jones, 2011). In addition, we now know that bacterial assemblages are composed of thousands of taxa with largely different environmental preferences, tolerances, and metabolism (Falkowski et al., 2008; Kirchman, 2016; Lennon, Aanderud, Lehmkuhl, & Schoolmaster, 2012). The latter insight has motivated much recent research on prokaryote distribution, diversity and dynamics across space and time, yet the factors and mechanisms responsible for that variability are not completely understood and

remain widely debated (Nemergut, Shade, & Violle, 2014; Prosser et al., 2007).

The primary goal of biogeography is to understand the distribution of organisms across space and time (Lomolino et al. 2010) and thus, the biogeography of bacteria includes both the description of the patterns of spatial and temporal variation in bacterial community structure as well as the exploration of the potential mechanisms or processes driving these patterns (Hanson, Fuhrman, Horner-Devine, & Martiny, 2012). A number of factors, acting at local and regional scales, can potentially interact to drive biogeographic patterns but in general, the differences in the composition between two or more bacterial communities are believed to result from two major biogeographic processes: 1) the dispersal of microorganisms between local communities, that determines the pool of bacteria that are transported to a given ecosystem and can potentially colonize it, as well as the rates at which taxa are transported and 2) sorting or selection of species from within that pool by local conditions, which will depend on bacterial ecophysiological features, environmental preferences and tolerances, and ecological strategies (Hanson et al., 2012; Lindström & Langenheder, 2012; Martiny et al., 2006; Székely, Berga, & Langenheder, 2012).

It has been hypothesized that, in general, compositional similarity between local bacterial communities should decrease with the degree of environmental heterogeneity but increase with the rates of dispersal of microbes between communities (Ramette & Tiedje, 2007; Ranjard et al., 2013; Villaescusa et al., 2010). Within dispersal-related processes, most studies have considered dispersal limitation, a historical process that can lead to divergence of very isolated bacterial communities (Martiny, Eisen, Penn, Allison, & Horner-Devine, 2011) and through which we would expect larger differences in composition between communities located far apart than between communities closer to each other. In highly connected systems or communities, however, mass effects (i.e., immigration of taxa from other ecosystems) can be important in determining the composition of local assemblages (Cottenie, 2005; Leibold et al., 2004). This process is expected to increase similarity between local communities, especially when the rates at which organisms are transported from one community to another are sufficiently high to lead to homogenization of the communities. As a consequence, low dispersal limitation or high mass effects would tend to generate more similar communities independently

of their environmental-related variation. However, these two processes usually operate at different spatial and temporal scales: dispersal limitation occurs between spatially separated communities or in the presence of physical barriers that imply large temporal scales, whereas mass effects will prevail when there is high connectivity between local assemblages at much higher rates (Lindström & Langenheder, 2012). On top of this complexity, depending on whether the environmental conditions are also spatially structured, like in the case of regions that are environmentally distinct, the resulting biogeographical patterns could be similar to those generated by the dispersal limitation scenario (Cottenie, 2005). All this suggests that a robust understanding of the mechanisms underlying the variation in bacterial community structure requires adopting approaches that consider the potential interactions between dispersal-related processes and environmental sorting across space and time.

So far, most studies on bacterial biogeography have revealed that bacteria consistently display biogeographic patterns and therefore that they are not randomly distributed (Hanson et al., 2012; Martiny et al., 2011). For example, the composition of bacterial communities is known to differ greatly between types of ecosystems (Barberán & Casamayor, 2010; Lozupone & Knight, 2007; Pommier et al., 2006) and within ecosystems at different temporal and spatial scales (Gilbert et al., 2012; Shade, Caporaso, Handelsman, Knight, & Fierer, 2013; Jones et al., 2012). Although the large majority of studies in bacterial biogeography often conclude that the environmental sorting is the dominant mechanism structuring bacterial communities across many different habitat types (Adams, Crump, & Kling, 2014; Langenheder & kely, 2011; Logares et al., 2013; Logue & Lindström, 2010; Van der Gucht et al., 2007), a growing body of evidence suggests that dispersal-related mechanisms can also play a role an important role in shaping this variability (Crump, Adams, Hobbie, & Kling, 2007; Crump, Amaral-Zettler, & Kling, 2012; Lindström, Forslund, Algesten, & Bergström, 2006; Nelson, Sadro, & Melack, 2009). However, a major challenge in contemporary microbial ecology is to disentangle the relative role of these different processes or mechanisms underlying the observed spatial and temporal patterns in bacterial taxonomic composition (Hanson et al. 2012), since obviously they can operate simultaneously.

0.2 The relevance of large-scale biogeography in the study of freshwater bacterioplankton diversity and ecology.

Freshwater ecosystems are one example where the influence of dispersal-related processes, in particular mass effects (i.e., the structuring of bacterial communities by massive advection of bacteria from other systems) can be particularly important because the flow of water through the landscape often results in a strong connectivity and intense exchange of microbes among local communities (Crump et al., 2007; 2012; Lindström et al., 2006; Nelson et al., 2009; Ruiz-González, Niño-García, & del Giorgio, 2015). However, most studies of freshwater bacteria continue to neglect this hydrologic dimension, and we know very little about how hydrology-driven processes may interact with local sorting of species to shape bacterioplankton communities (Crump et al., 2007; Lindström et al., 2006) for which we would need to cover an equally broad gradient in both environmental and hydrologic conditions. Rather, studies have typically targeted one or the other dimension, which has resulted in apparently contradictory results found across the literature: For example, those studies focusing on a single type of aquatic ecosystem (e.g., only lakes, only rivers) varying across spatial scales and environmental gradients (Ren et al., 2015; Yannarell & Triplett, 2005) generally conclude that environmental sorting by local conditions plays a major role in shaping local bacterial communities. On the other hand, studies that have attempted to cover wider gradients in hydrological conditions or landscape position, and thus include multiple ecosystems (e.g., streams, rivers, lakes), are typically restricted to small spatial scales (e.g., a single catchment), and thus usually conclude that mass effects are more important (Crump et al., 2007; 2012; Lindström et al., 2006; Nelson et al., 2009; Read et al., 2014).

Due to the directional movement of water from headwater streams to large rivers and lakes, one could expect that the relative importance of environmental sorting versus mass effects will likely vary depending on local hydrology and the position of the system along the hydrologic continuum: whereas the rate of water (and thus bacterial) inputs relative to ecosystem volume determines the intensity of mass effects, and the resulting water residence time in turn influences the extent of species sorting by bracketing

the time available for bacterial growth (Crump, Hopkinson, Sogin, & Hobbie, 2004; Lindström et al., 2006; Lindström & Bergström, 2004). Consequently, the fast movement of water in headwater ecosystems (i.e., small streams) and the associated mass effects should play a major role in determining bacterioplankton community composition in these systems, whereas local sorting of species will likely prevail in larger lakes and rivers with longer water residence time, where bacteria will have more time to respond to local conditions. This seems to explain the observed patterns of decreasing diversity and richness from headwater ecosystems towards downstream waters (Crump et al., 2012; Ruiz-González et al., 2015; Savio et al., 2014). Therefore, within a given catchment, the composition of bacterial communities should be determined by the interplay between hydrology, landscape position and local environmental conditions. However, over broader spatial scales (e.g. between different regions or catchments), bacterial community composition may further differ as a result of dispersal limitation or the increase in environmental ranges e.g.(Lear, Bellamy, Case, Lee, & Buckley, 2014; Liu, Yang, Yu, & Wilkinson, 2015), yet how hydrology and environmental conditions interact to shape bacterioplankton communities over large spatial scales remains completely unknown. Understanding the drivers and mechanisms shaping the structure of freshwater bacterioplankton communities thus requires incorporating the full range of variation in terms of ecosystem position, hydrologic regime, and environmental conditions within and across networks located in distinct regions, since studies covering only one axis of variability will likely provide a partial view of the actual complexity underlying the observed variation of bacterial community structure across the aquatic landscape.

0.3 Bacterial dominance and rarity: active and accidental taxa in freshwater bacterioplankton communities

This relevance of mass effects in shaping the structure of freshwater bacterioplankton communities will certainly condition the structure of bacterial assemblages, since it is possible that a large fraction of the taxa that are present are the result of hydrological transport and cannot develop in the recipient environment (Crump et al., 2007; Ruiz-González et al., 2015; Savio et al., 2014), and thus will not make part of the functional core of communities (Saunders, Albertsen, Vollertsen, & Nielsen, 2015). The advent of

high-throughput sequencing technologies has allowed the characterization of bacterial communities at an unprecedented level of resolution, leading to the detection of increasingly rare taxa (Pedrós-Alió, 2012). We know now that all bacterial communities are dominated by a few abundant taxa and an extremely large number of very rare bacteria, which has been referred to as the "rare bacterial biosphere" (Pedrós-Alió, 2006; 2012; Sogin et al., 2006), and which represents the most diverse component of microbial communities, comprising more that 90% of the taxa(Pedrós-Alió, 2012).

Abundant taxa within a given community tend to be reactive to temporal and spatial changes in environmental conditions and thus these taxa likely represent an active fraction of the community that is shaped through environmental sorting, and which dominates the functioning of communities. The extent to which this is also true for rare bacteria remains largely unknown, although the studies that have explored biogeographic patterns of both the dominant and the rare members of bacterial communities have often observed similar distribution patterns, concluding that they may all be structured by the same set of environmental factors (Campbell, Yu, Heidelberg, & Kirchman, 2011; Galand, Casamayor, Kirchman, & Lovejoy, 2009; Liu et al., 2015; Logares et al., 2013; Vergin, Done, Carlson, & Giovannoni, 2013). Accordingly, some recent studies have started to reveal that this rare biosphere, like its abundant counterpart, seems to be composed of bacteria displaying very different strategies, adaptations, preferences and origins (Campbell et al., 2011; Jones & Lennon, 2010; Lynch & Neufeld, 2015; Shade et al., 2014). For example, recent studies have proposed that some rare bacteria can be part of a "seed bank", a reservoir of dormant bacteria that can activate and grow under specific environmental conditions (Caporaso et al., 2012; Gibbons et al., 2013; Lennon & Jones, 2011). Others have shown that there are taxa that have intrinsically low growth and activity rates and for which rarity may be a survival strategy (Campbell et al., 2011; Logares, Mangot, & Massana, 2015; Lynch & Neufeld, 2015; Pedrós-Alió, 2012; Vergin et al., 2013). In other cases, however, the local low abundances of rare taxa have been attributed to an accidental presence driven by the transport of non-active or persistent taxa from other ecosystems (Hugoni, Taib, & Debroas, 2013; Pedrós-Alió, 2006; Vergin et al., 2013). A major unresolved question in contemporary microbial ecology concerns precisely the discrimination of the component of bacterial communities that results from passive transport and persistence, and which are therefore

mostly accidental, from the component that is reactive to local environmental and is therefore contributing to the functioning of the ecosystem. This question is particularly challenging in freshwater ecosystems, where, as mentioned previously, hydrologically mediated transport may enrich communities in this accidental component by carrying bacteria unable to thrive in the recipient environment.

So far, most studies have addressed the problem of activity of rare bacteria in natural communities by measuring the ratios of 16S rRNA to rRNA genes (rDNA) of individual bacterial taxa within a community in a limited amount of samples (Campbell et al., 2011; Hugoni et al., 2013; Jones & Lennon, 2010; Vergin et al., 2013), and by exploring shifts from rare to abundant of particular taxa over time (Alonso-Sáez, 2014; Alonso-Sáez, Díaz-Pérez, & Morán, 2015) or in response to experimental manipulations or natural gradients (Alonso-Sáez et al., 2015; 2014; Baltar et al., 2015). It is clear from these studies that at least some rare bacteria are active and that the boundary between the rare and dominant components of microbial communities is highly dynamic in time, yet they do not provide a broad perspective on how reactive versus accidental taxa are distributed across different ecosystems or what is their prevalence in a system over time (but see Shade et al. 2014), particularly within the rare biosphere.

In freshwater ecosystems, bacterial taxa that are not able to grow at some point in time or somewhere along a freshwater network will likely disappear as a result of dilution or death (Pedrós-Alió, 2012), and thus it is possible that the patterns of occurrence and abundance of individual freshwater bacterial taxa across space or over time reflect a combination of intrinsic eco-physiological features of taxa, which determine their potential growth rates, environmental breadth, and capacity to persist under unfavorable conditions, as well as of extrinsic factors related to their source, transport and dispersal. Since, as stated before, the relative influence of mass effects versus local sorting of species will change along the hydrologic continuum, the exploration of the patterns in abundance and distribution of taxa across an array of freshwater ecosystems varying in both environmental and hydrologic conditions (streams, rivers, lakes) may allow the identification of categories of bacterial SpAD. These categories may in turn allow separation of taxa whose spatial distribution is driven by mass effects from those whose distribution is driven by environmental sorting, regardless of their specific

environmental preferences. Most spatial studies in freshwater ecosystems so far have focused on the changes of bacterial communities across various ecosystems (Fierer & Jackson, 2006; Fierer, Morse, Berthrong, Bernhardt, & Jackson, 2007; Lozupone & Knight, 2007; Nemergut et al., 2011; Newton, Jones, Eiler, McMahon, & Bertilsson, 2011) showing recurrent biogeographic patterns at various taxonomic levels (Jones, Cadkin, Newton, & McMahon, 2012; Nemergut et al., 2011; Newton et al., 2011), taxon-specific differences in environmental breadth or responses (Barberán, 2014; Evans & Wallenstein, 2013; Fierer et al., 2007; Lennon et al., 2012; Newton et al., 2011) and clusters of taxa that share similar preferences based on their patterns of co-occurrence (Barberán, Bates, Casamayor, & Fierer, 2011; Comte, Lovejoy, Crevecoeur, & Vincent, 2016; Steele et al., 2011; Williams, Howe, & Hofmockel, 2014). However, whereas these studies have yielded valuable information regarding the environmental preferences of freshwater taxa, they provide little information regarding the actual assembly of these communities, and the extent to which the distribution of individual taxa is driven by key adaptive strategies or by passive, dispersal-related processes that control their local presence and abundance. Distinguishing between these two processes requires placing the patterns of SpAD not only in the context of environmental gradients, but also of the sources and movement of microbes across the landscape. Therefore, a quantification of the accidental and the reactive components of freshwater bacterioplankton communities will only be possible in the context of the whole aquatic network.

0.4 Objective of the thesis

The overall objective of this thesis is to improve our understanding of the large-scale drivers and mechanisms explaining the structuring of bacterioplankton communities across boreal lakes, taking into account the potential interactions between dispersal and mass effect processes associated with the directional movement of water within freshwater networks, and the role of species sorting by local conditions. In addition, the thesis uses patterns of spatial distribution of bacteria across boreal aquatic networks to discriminate between accidental taxa that are driven by passive transport along the network, from the active taxa that are subject to environmental selection within lakes,

and further explores the drivers underlying this partition. This general aim has been addressed by pursuing the following three specific objectives, each representing one chapter of the present thesis:

1. Understand the main underlying mechanisms and drivers of large-scale biogeographic patterns of bacterioplankton community structure across boreal lakes of Quebec, which differ widely in their environmental conditions and geographic position.
2. Identify the core active and the passive accidental components within these boreal lake bacterioplankton communities, based on their large-scale spatial distribution patterns, and explore their ecological characteristics.
3. Explore the temporal dynamics of these spatially-derived active and accidental categories of bacterial taxa, in order to determine whether their associated properties are maintained over time and are thus intrinsic features of taxa.

0.5 General Approach

The accomplishment of the overall objectives of this thesis relies on a sampling design that allows capture of the variability in taxonomic structure of lake bacterial communities over a broad range of environmental, hydrological and landscape conditions, as well as over space and time, in order to elucidate the interactions between the different biogeographic processes and drivers shaping the taxonomic structure of bacterioplankton communities. To do so, we conducted an unprecedented, large-scale spatial survey of over 380 boreal lakes and rivers, whose bacterioplankton communities were sampled and characterized using Illumina sequencing of the 16S rRNA gene, only once in summer (chapter I and II). We also carried out a temporal survey in which bacterioplankton communities from a single lake were sampled monthly over an annual cycle, and 21 lakes were sampled on three occasions (spring, summer and fall - chapter III-). In addition, to establish the links between the relative importance of mass effects versus environmental sorting on the spatial-temporal distributions of accidental vs. active categories of taxa, in chapter II we first grouped individual OTUs into four distinct categories of spatial abundance distribution (SpAD), by fitting their

abundance distributions across lakes to several statistical model distributions (normal-like, bimodal, logistic, lognormal). Then, we assessed the ecological underpinnings of the resulting categories by exploring the responses of these taxa to environmental gradients covered by our study lakes and by retracing the behavior of each OTU within the fluvial networks associated with these lakes. In order to test whether specific ecological properties derived from spatial patterns of taxa were also maintained over time, in chapter III we analyzed the temporal dynamics of the OTUs that had been previously grouped within the categories of SpAD in (chapter II) along an annual cycle in the oligotrophic lake Croche (Québec, Canada), and we assessed the coherence between the ecological properties of the OTUs within each spatial category over space and time. To extend these results beyond lake Croche, we further compared the spatial and temporal dynamics of the mentioned OTUs across 21 additional boreal lakes for which we had data on three occasions (spring, summer and fall).

0.5.1 Spatial survey (Chapters I and II)

The objective for chapters I and II was to maximize the environmental, hydrologic and geographic gradients covered in our study so as to be able to detect patterns in taxonomic composition related to dispersal limitation and mass effects. For this, we sampled 198 lakes located in seven different regions of boreal Quebec between 2009 and 2013. These regions (Laurentides, Abitibi, James Bay, Saguenay, Chibougamau, Schefferville, La Côte-Nord, Figure 1.1a) cover a total area of ca. 900,000 km^2 (44-56°N, 64-80°W) and span large gradients in climate, limnological properties, land cover, and geomorphology typical of the boreal zone. This sampling regimen resulted in a set of lakes exposed to gradual differences in mean annual temperature and precipitation, different dominant vegetation, human perturbation of the landscape, as well as displaying random variability in catchments, lakes size (0.002-4345 km^2), trophic status, limnological conditions, and distances between them.

In addition, since boreal lakes are rarely isolated ecosystems, but are rather embedded in complex aquatic networks, and are connected with other lakes through the movement

of water in the landscape, understanding the mechanisms shaping lake bacterioplankton community structure necessarily requires placing lakes in the context of these aquatic networks. Thus, we further included in our analysis 188 streams and rivers from the same regions where the lakes were sampled, which covered a broad range of variation in river order (Strahler order 0-8), hydrologic regime, network position and environmental conditions. All these sites were sampled only once during the summer (July/August) to minimize seasonal variation, and all the sites in each region were sampled within the same year. Water samples were taken at a depth of 0.5 m at the deepest measured spot of lakes and near the shore in rivers.

0.5.2 Temporal survey (Chapter III)

To address the specific objective of Chapter III, we carried out a seasonal study in lake Croche, located in the Laurentian region of Québec (45.99°N 74.00°W), which was sampled monthly from July 2012 to October 2013. We collected epilimnetic (0.5-1 m depth) water samples, and measured *in situ* several physicochemical parameters. During winter, sampling was carried out through the ice. Moreover, during 2012 and 2013, a subset of 21 lakes from four different boreal regions in Quebec (Abitibi, Bay James, Chibougamau, Saguenay) used in Chapters I and II were also sampled in spring and fall. This sampling allowed us to explore the temporal consistency of the patterns found at the spatial scale, and to gain insight into temporal patterns that may be neglected considering that our spatial survey represented only a summer snap-shot of bacterioplankton communities.

0.5.3 Environmental, hydrologic and landscape variables

To characterize environmental and limnological heterogeneity across the dataset, we measured key environmental variables *in situ*, such as temperature, dissolved oxygen, pH, and conductivity, and we took water samples for laboratory measurements of

Chlorophyll *a* (Chl*a*) concentration, DOC concentration and composition, total phosphorus (TP) and nitrogen (TN) concentrations. In addition, we also characterized regional and hydrologic variability by calculating several geographic variables, such as lake area, river length and order, catchment areas, the area covered by water in the catchment, water residence times, mean annual temperature, precipitation, runoff and elevation of the sampled sites. A summary of the ranges of the main variables used for each chapter can be seen in Table 1.1 and Supplementary Table 2.1 and 3.1.

0.5.5 Characterization of bacterial taxonomic community composition

To characterize bacterial community composition for all the mentioned sites, we filtered 300-500 ml water samples onto 0.22µm pore-size filters for further genomic DNA extraction. DNA was extracted using the MoBio PowerWater DNA extraction kit following the manufacturer's specifications. Libraries of the V3-V4 region of the 16SrRNA were made using the primers 515F and 806R, and sequenced on an Illumina MiSeq2000 following a paired-end approach (Caporaso *et al.*, 2012). Paired-end reads were assembled with FLASH (Magoč & Salzberg, 2011) and sequences between 250-290 bp were used for downstream analyses in QIIME to remove primers and low-quality, archaeal, and chloroplast reads (Caporaso *et al.*, 2010). After the detection and removal of chimeric sequences, quality sequences were clustered into operational taxonomic (OTUs). Whereas for Chapter I sequences were clustered into OTUs ($\geq$ 97% sequence similarity) using UCLUST (Edgar, 2010), for Chapter II and III we used a recently developed clustering method that avoids using arbitrary thresholds of similarity, the SWARM 2.0 algorithm (Mahé, Rognes, Quince, de Vargas, & Dunthorn, 2014). The two approaches yielded almost exactly similar patterns regarding the composition of bacterial communities, yet the SWARM algorithm resulted in a higher percentage of taxonomically assigned sequences relative to UCLUST and the clustering process was many times faster. We further assigned taxonomy to the representative sequences by using SILVA 111 reference database (Quast et al., 2012) and the RDP classifier (Wang, Garrity, Tiedje, & Cole, 2007). To enable comparisons between samples, the OTU tables were randomly subsampled to ensure an equal number of sequences per sample,

based on the sample with the least number of reads.

0.5.6 Statistical analyses

We used nonmetric multidimensional scaling (Vegan metaNMDS, Oksanen *et al.* 2013) with Bray-Curtis distances to analyze spatial patterns in community composition, and the Vegan envfit function (Oksanen *et al.* 2013) to explore the factors explaining the variation of these patterns. Differences in environmental and taxonomic composition among categories (i.e., regions, ecosystems) were tested using Permanova with 999 permutations (Anderson, 2011), (Anderson, 2006). To estimate the contribution of different factors in explaining the variation of the large-scale patterns of bacterial community composition, we performed a variation partitioning analysis using the NMDS scores as response variables (Borcard, Gillet, & Legendre, 2011). We also used Mantel tests (R Vegan Package, Okasanen et al. 2015) to calculate the correlation between distance and community dissimilarity matrices. Spatial and temporal OTUs turnover (Chapters II and III) were estimated using the beta.pair function, as the turnover-fraction of Jaccard pair-wise dissimilarity (R betapart package, (Baselga & Orme, 2012). All the analyses were made using the R 3.0.0 software (R Core Team, 2013).

0.6 Thesis structure

The results of this thesis are presented in three chapters, one of which is already published (Chapter I). Chapter II was recently submitted, and Chapter III is currently in preparation for submission.

Chapter I. Niño-García, J.P., Ruiz-González, C. & del Giorgio, P.A. 2016. Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. The ISME Journal, 1755-1766.

Chapter II. Niño-García, J.P., Ruiz-González, C. & del Giorgio, P.A. 2016. Landscape-

scale spatial abundance distributions discriminate core from random components of boreal lake bacterioplankton. Ecology Letters, 1506-1515.

Chapter III. Niño-García, J.P., Ruiz-González, C. & del Giorgio, P.A. Exploring the ecological coherence between the spatial and temporal patterns of bacterioplankton in boreal lakes. Submitted to Frontiers in Microbiology.

CHAPTER I

INTERACTIONS BETWEEN HYDROLOGY AND WATER CHEMISTRY SHAPE BACTERIOPLANKTON BIOGEOGRAPHY ACROSS BOREAL FRESHWATER NETWORKS

Juan Pablo Niño-García[1,2], Clara Ruiz-González[1], Paul A. del Giorgio[1]

Published in *ISME JOURNAL* (2016): 1755-1766

[1]Groupe de recherche interuniversitaire en limnologie, Département des Sciences Biologiques, Université du Québec à Montréal, CP 8888, succursale Centre-Ville, Montréal, QC, H3C 3P8, Canada

[2]Escuela de Microbiología, Universidad de Antioquia, Ciudad Universitaria Calle 67 No 53-108, Medellín, Colombia

N.B References cited in this chapter are presented at the end of the thesis.

1.1 Abstract

Disentangling the mechanisms shaping bacterioplankton communities across freshwater ecosystems requires considering a hydrologic dimension that can influence both dispersal and local sorting, but how the environment and hydrology interact to shape the biogeography of freshwater bacterioplankton over large spatial scales remains unexplored. Using Illumina sequencing of the 16S rRNA gene, we investigate the large-scale spatial patterns of bacterioplankton across 386 freshwater systems from seven distinct regions in boreal Québec. We show that both hydrology and local water chemistry (mostly pH) interact to shape a sequential structuring of communities from highly diverse assemblages in headwater streams towards larger rivers and lakes dominated by fewer taxa. Increases in water residence time along the hydrologic continuum were accompanied by major losses of bacterial richness and by an increased differentiation of communities driven by local conditions (pH and other related variables). This suggests that hydrology and network position modulate the relative role of environmental sorting and mass effects on community assembly by determining both the time frame for bacterial growth and the composition of the immigrant pool. The apparent low dispersal limitation (i.e., the lack of influence of geographic distance on the spatial patterns observed at the taxonomic resolution used) suggests that these boreal bacterioplankton communities derive from a shared bacterial pool that enters the networks through the smallest streams, largely dominated by mass effects, and that is increasingly subjected to local sorting of species during transit along the hydrologic continuum.

1.2 Introduction

The widespread use of high-throughput sequencing approaches in microbial ecology studies is resulting in unprecedented insight into the structure of prokaryotic communities. We now know that the taxonomic composition of bacterial communities differs greatly between types of ecosystems (Barberán & Casamayor, 2010; Lozupone & Knight, 2007; Pommier et al., 2006) and within ecosystems at different temporal and spatial scales (Gilbert et al., 2012; Shade, Caporaso, Handelsman, Knight,

& Fierer, 2013; Jones, Newton & McMahon, 2012). In general these differences are driven by two major mechanisms: 1) dispersal of microorganisms among communities, which determines the bacteria that can be transported and potentially colonize a given ecosystem, and 2) selection of species (i.e. species sorting) from within this immigrant pool on the basis of their environmental preferences and local conditions (Hanson, Fuhrman, Horner-Devine, & Martiny, 2012; Lindström & Langenheder, 2012; Martiny et al., 2006; Székely, Berga, & Langenheder, 2012).

Although there is ample evidence supporting a major role of environmental sorting in shaping bacterial community structure across many different habitats types (Adams, Crump, & Kling, 2014; Langenheder & kely, 2011; Logares et al., 2013; Logue & Lindström, 2010; Van der Gucht et al., 2007), in freshwater ecosystems the influence of mass effects (i.e., the structuring of bacterial communities by massive advection of bacteria from other systems) can be particularly important because the flow of water through the landscape often results in strong connectivity and intense exchange of microbes among local communities (Crump, Adams, Hobbie, & Kling, 2007; Crump, Amaral-Zettler, & Kling, 2012; Lindström, Forslund, Algesten, & Bergström, 2006; Nelson, Sadro, & Melack, 2009; Ruiz-González, Niño-Garcia & del Giorgio, 2015b). It is thus likely that both the environmental and hydrological conditions interact to shape the patterns that we observe in freshwater bacterioplankton communities, but how this interaction influences the biogeography of bacterioplankton over large spatial scales remains largely unexplored.

At the catchment scale, the movement of water within the network has been shown to play a key role in modulating the relative importance of species sorting and mass effects in freshwater bacterioplankton community assembly: whereas the rate of water (and thus bacterial) inputs relative to the ecosystem volume determines the intensity of the mass effects, the resulting water residence time in turn influences the extent of species sorting by local conditions by bracketing the time available for bacterial growth (Crump, Hopkinson, Sogin, & Hobbie, 2004; Lindström et al., 2006; Lindström & Bergström, 2004). In addition, because water and bacteria move directionally, most ecosystems act both as recipients and sources of bacteria, and the immigrant pool arriving to a given ecosystem will also vary depending

on its position along the hydrologic continuum (Nelson *et al.*, 2009). The local bacterioplankton community composition within a given aquatic network will thus be determined by the hydrology, the position of the system in the network, and the local environmental conditions, and consequently, the relative importance of the community assembly processes will likely vary along the hydrologic continuum.

Beyond individual catchments and over broader spatial scales, however, bacterial taxonomic composition may further differ between communities as a result of increases in environmental ranges relative to the local conditions, but also due to dispersal limitation among different geographic regions (for example, Lear *et al.*, 2013; Liu *et al.*, 2015), which will determine the composition of the regional species pool. Thus, the large-scale spatial patterns in composition of freshwater bacterioplankton communities likely result from combinations of these local and regional processes, as conceptualized in Figure 1.1. For example, we may detect differences in the composition of bacterial communities inhabiting ecosystems with long water residence times (for example, lakes) either because they are located in areas that do not share the same regional bacterial pool due to low dispersal (Fig. 1.1a), or because their environmental differences may have selected different taxa from a shared pool of bacteria (Fig. 1.1b). Similarly, communities subjected to strong mass effects, (e.g inhabiting headwater streams with short residence times) may show differences in composition if influenced by distinct bacterial pools (Fig. 1.1a), but may be compositionally similar if derived from a common immigrant pool (Fig. 1.1b), regardless of their local environmental conditions. An important corollary of this conceptual framework is that disentangling the major processes underlying the spatial patterns in bacterial community composition requires incorporating the full range of variation in terms of ecosystem position, hydrologic regime, and environmental conditions within and across networks located in distinct regions.

Most studies to date, however, have covered only limited portions of this multidimensional space. For example, the literature abounds in studies focusing on a single type of aquatic ecosystem (for example, only lakes, only rivers, Fig. 1.1-scenarios 2 and 3) varying across spatial scales and environmental gradients (Ren et al., 2015; Yannarell & Triplett, 2005). These studies typically conclude that local conditions play a major role in shaping local community composition through species

sorting. On the other hand, studies that have included multiple ecosystems (for example, streams, rivers, lakes, Fig. 1.1- scenario 1), which may cover wide ranges of network positioning and hydrologic conditions but that are restricted to small spatial scales (for example, a single catchment, Lindström *et al.*, 2006; Crump *et al.*, 2007; Nelson *et al.*, 2009; Crump *et al.*, 2012; Read *et al.*, 2015), often conclude that hydrology and the associated mass effects are major drivers of community assembly. The framework presented in Figure 1.1 suggests that the ambiguous and sometimes contradictory conclusions found in the literature concerning the main drivers of the spatial variability in aquatic bacterial community composition are neither incorrect nor mutually exclusive, but rather represent fragments of a more complex biogeographic reality. Understanding the mechanisms underlying this large-scale biogeography thus requires a cross-regional, whole network perspective that has seldom, if ever, been undertaken.

Here we explore the large-scale spatial patterns in freshwater bacterioplankton community composition across the boreal biome of Québec, and we infer its main underlying drivers and mechanisms by exploring the variability in bacterial taxonomic structure over broad ranges in hydrologic and environmental conditions, network position, and geographic distance. We have characterized the taxonomic composition of 386 bacterioplankton communities inhabiting freshwater ecosystems positioned along the entire aquatic continuum, ranging from the smallest headwater streams to the largest rivers and lakes, which belong to seven different geographic regions that span large gradients in environmental, climate, and landscape properties. This sampling design allowed us to understand how local environmental and hydrologic conditions, network position, and dispersal-related regional differences interact to shape bacterial community composition across these complex boreal aquatic networks.

1.3 Materials and methods

1.3.1 Study site and experimental design

Water samples were collected from 198 lakes and 188 rivers across seven boreal regions of Québec (Canada) between 2009 and 2013. The regions sampled (Laurentides, Abitibi,

James Bay, Saguenay, Chibougamau, Schefferville, La Côte-Nord, Fig. 1.2a) cover a total area of ca. 900,000 km² (44-56°N, 64-80°W) and span large gradients in terms of climate, limnological properties, land cover, and geomorphology typical of the boreal zone (for details see Lapierre *et.al,* 2015, Rasilo *et al.* 2015, Ruiz-González *et al.* 2015a). The studied sites covered the entire range in lake area (0.002-4345 km²) and river order (Strahler order 0-8), and were sampled once during the summer (July/August). All the sites in each region were sampled within the same year. Samples were taken at a depth of 0.5 m at the deepest measured spot of lakes and near the shore in rivers. Temperature, dissolved oxygen, pH, and conductivity were measured in situ with a YSI probe. Samples were filtered in situ through 0.45 µm and stored in acid-washed glass vials for DOC and optical analyses, or kept in the dark in acid-rinsed bottles for further processing.

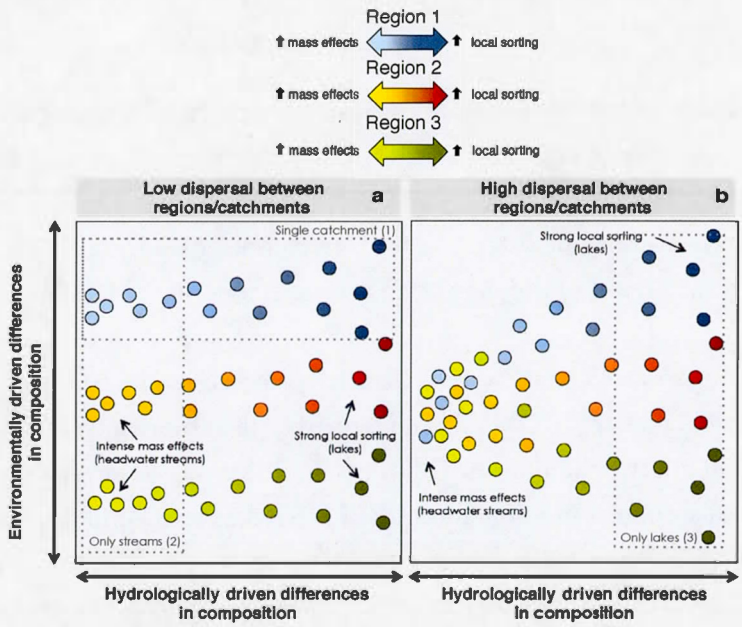


Figure 1.1. Schematic representation of potential large-scale spatial patterns in bacterioplankton taxonomic composition across freshwater networks, where dots represent individual bacterioplankton communities in an ordination space, based on their compositional dissimilarity, and colors indicate the different catchments or geographic regions to which they belong. Due to the movement of the water in the landscape, the composition of local bacterioplankton communities within a single catchment or region will differ depending on their position along the hydrologic continuum (horizontal axis), since they will be differentially affected by hydrology that regulates the mass effects versus environmental sorting ratio (see Introduction for further explanation), represented here by the light to dark color gradient. As a result, the compositional variation due to local conditions (vertical axis) will likely increase along the hydrologic continuum due to a gradual intensification of the local sorting of species with increasing water residence time. Over broader spatial scales, the increase in environmental differences will likely result in stronger environmentally-driven dissimilarities between communities (i.e., along the vertical axis), but the overall large-scale spatial patterns may further differ depending on the degree of dispersal of microbes between regions or catchments: For example, in a case of low dispersal between regions (a), a regional structuring of the communities may be detected even in systems with little or no

environmental sorting (i.e. headwater streams) because the regional bacterial pools are different due to dispersal limitation. In contrast, under a scenario of high dispersal between regions (b), a common bacterial pool will lead to little environmentally driven compositional differences in the headwaters but to a increasing differentiation of communities towards downstream systems (i.e., lakes) due to stronger local sorting of species. From this perspective, whereas communities located at the headwaters of the aquatic continuum will mostly reflect the nature of the regional bacterial pools due to strong mass effects, communities located further downstream in the network will be the result of species sorting by the local aquatic conditions. The interpretation of the factors and mechanisms underlying the observed spatial patterns in bacterial community composition will thus depend on the portion of this complex space that is considered (dotted areas, scenarios 1, 2 and 3). For example, studies based on a narrow spatial scale, as depicted in scenario 1, might conclude that hydrology is the main driver of changes in bacterioplankton community composition. In contrast, studies that cover broader environmental gradients but are limited to certain portions of the hydrologic continuum (scenarios 2 and 3) may conclude that either dispersal limitation (2) or local sorting (3) shape local community assembly. We argue that a cross-regional, whole network approach is necessary to disentangle the influence of the mechanisms and factors that are actually influencing the local assembly of bacterioplankton communities in complex freshwater networks.

1.3.2 Chlorophyll, chemical, and optical analyses

Chlorophyll *a* (Chl*a*) concentration was determined spectrophotochemically in hot ethanol extracts (90%). DOC concentration was measured on an OI 1010 TOC analyzer. Total phosphorus (TP) and nitrogen (TN) concentrations were analyzed after persulfate and alkaline persulfate digestions, respectively, following standard methods (see Rasilo *et al.* 2015).

The optical properties of DOC were measured as indices of its composition. Colored dissolved organic matter (cDOM) was quantified as the absorbance at 440nm using an Ultrospec 3100 spectrophotometer. DOC composition was described on the basis of fluorescence absorption/emission spectra (EEMS), measured in a Shimadzu RF5301 PC spectrofluorophotometer across excitation/emission wavelengths of 275–450 nm and 280–600 nm, respectively. Six main fluorescence components related either with refractory-humic (C1-C3) or with biolabile, freshly-produced DOM (C4-C6) material were recovered from the EEMS using parallel factor analysis (PARAFAC, for details see Lapierre and del Giorgio 2014). The percent contribution of each component was calculated relative to the total fluorescence of the six PARAFAC components.

1.3.3 Geographic analyses

Lake areas, river length and order, catchment areas, the area covered by water in the catchment, and elevation of the sampled sites were derived using the ArcMap 10 and ArcGIS V10 software (ESRI Inc., Redland, CA) applied on the DEM derived from (1:50000) maps. For each catchment area, we averaged catchment slope based on DEM, as well as various land cover properties obtained from Geobase (2009). Mean annual temperature, precipitation, and runoff were extracted at each site location from a long-term climate database (WordClim, Hijmans et al. 2005). The water residence time (WRT) of lakes was estimated from lake volume (estimated as mean depth x lake area), catchment area, and mean annual regional runoff. River WRT was calculated based on the measured water velocity at the sampling point and the total upstream distance.

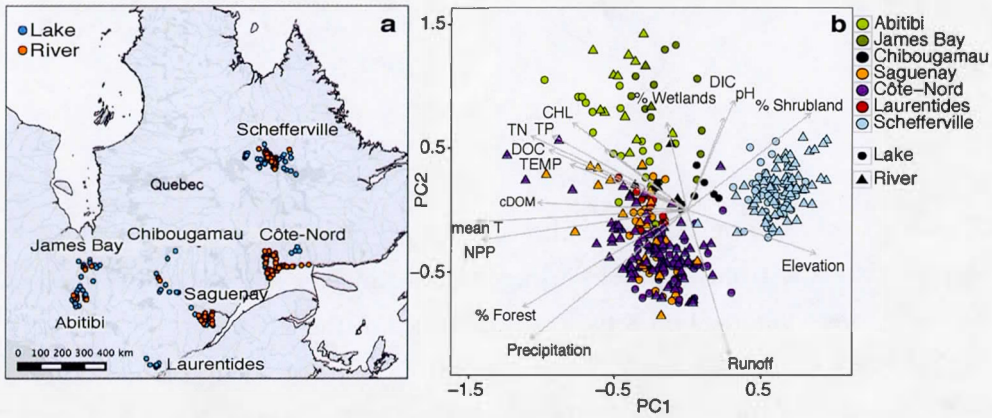


Figure 1.2. A) Distribution of sampling sites across the seven sampled regions in Northern Quebec (Canada). B) Principal component analysis (PCA) of the sites based on the measured environmental and geographic parameters. Different regions are indicated by different colors and symbols indicate rivers (triangle) or lakes (circles). The two first axes explain 44.5% of the variance. [DOC] Dissolved organic carbon; [DIC] dissolved inorganic carbon; [cDOM] colored dissolved organic matter; [%C2] percentage of fluorescent component C2 (humic-like DOM); [%C5] percentage of fluorescent component C5 (freshly produced labile DOM); [TP;TN] Total phosphorus and nitrogen; [Chla] Chlorophyll a;[Precipitation] Mean annual precipitation; [Mean T] Mean annual temperature; [Runoff] Mean annual runoff; [Elevation] mean catchment elevation; [%Forest/%Shrubland/%Wetland] % forest/shrubland/wetland covered area in the catchment; [Temp] Water temperature; [Cond] Conductivity; [NPP] Net primary productivity.

1.3.4 Water isotopes

Samples were collected to determine ^{18}O and ^{2}H isotopes by laser spectroscopic analysis of liquid water technique (LGR DT-100 Liquid Water Stable Isotope Analyzer, Los Gatos Research Inc., Mountain View, CA). We used water isotopic composition to calculate the deuterium excess (d-excess), as: d-excess = $\delta^{2}H - 8 \times \delta^{18}O$

Which provides an index of the degree of evaporation relative to precipitation, and therefore, the relative proximity to groundwater sources. Since headwater systems are strongly influenced by soil groundwater, the water carries an isotopic signature that closely resembles that of precipitation (high d-excess). Given that temperature increases and d-excess decreases as water transits through the network and is subjected to warming and evaporation (Gibson, Prepas, & McEachern, 2002; Turner, Edwards, & Wolfe, 2014), here we use water temperature and d-excess as proxies of the groundwater-mediated connectivity with the surrounding catchment.

1.3.5 Bacterial community composition

In all, 300-500 ml water samples were filtered onto 0.22μm pore-size filters, and genomic DNA was extracted from the filters using the MoBio PowerWater DNA extraction kit following the manufacturer's protocol. Libraries of the V3-V4 region of the 16S ribosomal RNA were made using the primers 515F and 806R, and sequenced on an Illumina MiSeq2000 following a paired-end approach (Caporaso et al., 2012). Paired-end reads were assembled with FLASH (Magoč & Salzberg, 2011) and sequences between 250-290 bp were used for downstream analyses in QIIME to remove primers and low-quality, archaeal, and chloroplast reads (Caporaso et al., 2010). After the detection and removal of chimeric sequences with ChimeraSlayer (Haas et al., 2011), quality sequences were aligned with Mothur aligner (Schloss et al., 2009) and clustered into operational taxonomic units (OTUs, ≥ 97% similarity) using UCLUST (Edgar 2010). In order to assign the taxonomy to the representative sequences, we used the

RDP classifier and the Ribosomal Database Project taxonomy (Wang, Garrity, Tiedje, & Cole, 2007). We discarded all OTUs that were represented in less than 10 samples and/or by less than 10 sequences. To enable comparisons between samples, the OTU table was randomly subsampled to ensure an equal number of sequences per sample, based on the sample with the least number of reads (50,323 sequences).

1.3.6 Statistical analyses

We used nonmetric multidimensional scaling (Vegan metaNMDS, Oksanen et al. 2015) with Bray-Curtis distances to ordinate the samples based on their dissimilarity in community composition, and the Vegan envfit function (Oksanen et al. 2015) to fit environmental vectors onto the ordination space. Differences in environmental and taxonomic composition among categories (i.e., regions, ecosystems) were tested using Permanova with 999 permutations (Anderson, 2006, 2011). To estimate the relative contribution of environmental and hydrological factors in explaining the variation of the large-scale patterns of bacterial community composition, we performed a variation partitioning analysis using the NMDS scores as the response variables. We grouped the variables as environmental or hydrological factors, which, after a forward selection procedure, were used as the explanatory groups of variables (Borcard, Gillet, & Legendre, 2011). All the analyses were made using the R 3.0.0 software (R Core Team, 2013).

1.4 Results

1.4.1 Large-scale patterns in bacterial community composition across boreal aquatic ecosystems.

The studied lakes and rivers segregated into clearly differentiated groups of sites based on their environmental, climatic, and landscape properties (Fig. 1.2a,b). These groups roughly corresponded to the different geographic regions (PermanovaBY_ REGION R^2=0.43, $p<0.01$), although systems located in Laurentians, Côte-Nord

and Saguenay tended to overlap. Sites were segregated among these regional groups based on differences in water chemistry (pH, DOC, DIC and cDOM) and trophic status (TP, TN and Chl*a*), and lakes and rivers from a given region tended to cluster together. Briefly, the sites varied along a gradient from alkaline and more productive waters in Abitibi and Bay-James to acidic and less productive waters in Saguenay and Côte-Nord. These environmental gradients were also related with changes in mean annual temperature and regional net primary productivity (as illustrated by the PC1 axis) as well as mean annual precipitation and runoff (PC2).

This environmental segregation of sites was not as clearly reflected in terms of the composition of bacterial communities, and the NMDS analysis showed much smaller differences between the geographical regions on the basis of bacterial composition (Fig. 1.3a, PermanovaBY_REGION R^2=0.10, $p<0.01$), and a significant overlap in composition between sites from different regions (Fig. 1.3a). River and lake communities, in contrast, showed a clearer segregation, although there was some degree of overlap between them (Fig. 1.3, PermanovaBY_ ECOSYSTEM R^2=0.11, $p<0.01$) since the largest rivers (Strahler order>4) clustered together with lakes (details not shown). Interestingly, river communities were much more dispersed along the NMDS1 axis than lake communities (Fig. 1.3).

1.4.2 Drivers of bacterial community composition across freshwater ecosystems.

In order to identify the drivers explaining the spatial patterns in community composition depicted by the NMDS in Figure 1.3, we fit the measured environmental and landscape variables onto the ordination. pH, water retention time (WRT), and water temperature were the variables most strongly correlated to the two NMDS axes (Fig. 1.3, Table 1.1). NMDS1 was strongly correlated to hydrologic and landscape variables such as water temperature, WRT, d-excess, catchment area, and the percentage of water in the catchment. This NMDS axis also represented to some extent network positioning, with scores declining with increasing river order along the hydrologic continuum, and also with increasing distance to the headwaters, thus separating small headwater streams from rivers, and lumping together the largest rivers with lakes (details not

shown). NMDS2 was mostly related to water chemistry, and particularly to pH (but also to cDOM, DIC, and specific DOM fluorescence components, Fig. 1.3, Table 1.1).

The geographic location (latitude and longitude) was not significantly related to either of the NMDS axes, and there was no significant relationship between the spatial distance and the taxonomic dissimilarity between sites (Partial Mantel R=0.06 p=0.1, Fig. 1.4a), thus indicating that the observed regional segregation described above (Fig. 1.3a) was not primarily driven by dispersal limitation. Rather, the community dissimilarity was significantly correlated to both the environmental and hydrologic distances between pairs of sites (Fig. 1.4b and c, respectively). Environmental and hydrologic distances were only weakly correlated to each other (details not shown), such that any given range of hydrologic conditions was associated to the entire range in environmental variability, suggesting that our sampling scheme captured the full range in both dimensions that exists across the boreal biome.

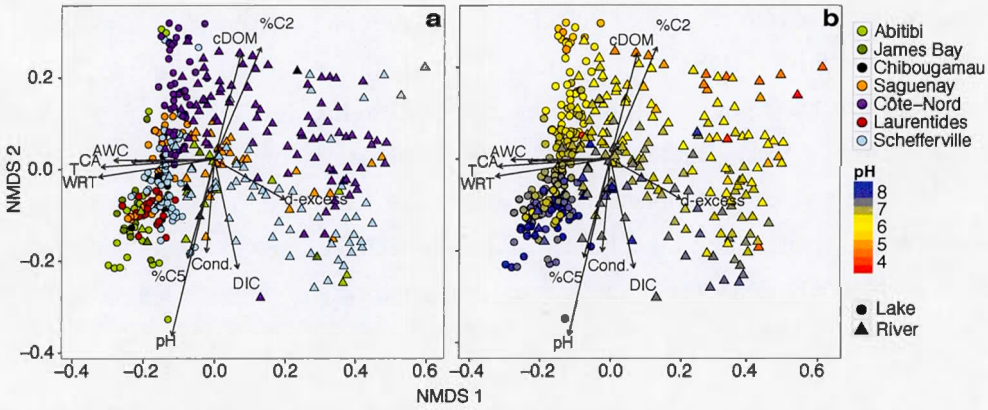


Figure 1.3. NMDS ordination of bacterioplankton communities based on the Bray-Curtis dissimilarity of community composition (stress=0.13). Shape indicate rivers (triangles) or lakes (circles), and sites are colored according to geographical region (a) and water pH (b), which was the variable that best fitted the ordination space (R^2=0.61, see Table 1). The arrows indicate the direction at which the environmental and hydrological vectors fit the best (using envfit function) onto the NMDS ordination space. The size of the arrow is proportional to the strength of the correlation of each variable. [T] Water temperature; [CA] Catchment area; [AWC] area of water in the catchment; [WRT] Water residence time; [DIC] dissolved inorganic carbon; [cDOM] colored dissolved organic matter; [%C2] percentage of fluorescent component C2; [%C5] percentage of fluorescent component C5 (freshly-produced labile DOM); [Cond] conductivity.

1.4.3 Changes in taxonomic diversity and composition along the hydrologic continuum.

In order to further understand the observed changes in community composition along the hydrologic continuum, we divided the sampled sites into seven groups (G1-G7) that were equally spaced along the NMDS1 axis. These groups were associated to different average network position and water retention time within the hydrologic continuum: Most lakes were contained within groups G1 and G2, the smallest headwater streams were mostly included in G7, and the remainder of the rivers and some lakes were distributed among the other groups. For each of the seven groups we calculated the average OTU richness, average Pielou's evenness index, and the percentage of sequences within the major bacterial Phyla. This analysis revealed that sites with the highest scores on the NMDS1 axis (G7, smallest streams) showed the highest average OTU richness (Fig. 1.5a) and evenness (Fig. 1.5b), and that there was a striking decline in both indices towards sites with the lowest average NMDS1 scores (G1, larger rivers and lakes, Fig. 1.5a,b): on average, 6000 OTUs per community were lost between these two groups of sites, and the evenness decreased from 0.86 to 0.66. The taxonomic composition also changed gradually along this NMDS1 axis, resulting in highly differentiated communities at the extremes of this hydrologic continuum even at the Phylum level (Fig. 1.5c). Overall, whereas Actinobacteria, Bacteroidetes, and Cyanobacteria increased towards larger rivers and lakes, Betaproteobacteria increased towards larger rivers but decreased again in lakes. The remaining classes of Proteobacteria and dominant soil groups such as Acidobacteria systematically decreased from the smallest streams to lakes.

1.4.4 Shifts in the relative influence of the environment versus hydrology on bacterial community composition.

The pronounced loss of OTUs downstream the hydrologic continuum, together with the decline in taxonomic evenness associated to the dominance of certain bacterial groups (Fig. 1.5c, Supplementary Fig. S1.1), suggest that the relative importance of mass effects and environmental sorting of species may change along the hydrologic continuum. To

test this hypothesis, we calculated the correlation between taxonomic and environmental dissimilarity matrices for each of the seven groups of sites along the NMDS1 axis (G1 to G7, described above, Fig. 1.6a). We used differences in pH as a proxy of environmental distance because it was the variable most strongly correlated to changes in taxonomic composition along the NMDS2 axis (Fig. 1.3, Table 1.1). The correlation between the environmental and the compositional dissimilarities between sites were weakest in the groups with the shortest average water retention times and highest position within the network (i.e. smallest headwater streams, G7, Mantel R=0.25), and increased gradually towards larger rivers and lakes, where changes in pH explained a much larger portion of the variability in taxonomic composition (G1, Mantel R=0.49, Fig. 1.6a).

	NMDS1	NMDS2	R^2	p	Range
Temp (°C)	-0.99881	-0.0488	0.401	0.001	6.24 - 32.3
WRT (days)	-0.99124	-0.13205	0.4185	0.001	0.01 - 10523
WC (%) Km^2	-0.99853	-0.05429	0.3162	0.001	0.002 - 12753
CA (Km^2)	-0.99484	-0.10143	0.2177	0.001	16 - 55148
d-excess	0.92625	-0.3769	0.1889	0.001	-5.44 - 37
pH	-0.29484	-0.95555	0.6178	0.001	3.3 - 8.95
cDOM	0.31115	0.95036	0.2318	0.001	0.01 - 58.8
DIC	0.29063	-0.95684	0.2285	0.001	0.282 - 42.2
%C2	0.48359	0.8753	0.2873	0.001	0.13 - 0.48
%C5	-0.32545	-0.94556	0.2047	0.001	0.05 - 0.25

Table 1.1 Envfit results and range of the environmental and hydrological variables included in the analysis shown in Figure 1.3. The degree of fit within the whole ordination space (R2) and for each ordination axis (NMDS1, NMDS2) are shown. [T] Water temperature; [CA] Catchment area; [WRT] Water residence time; [DIC] dissolved inorganic carbon; [cDOM] colored dissolved organic matter; [%C2] percentage of fluorescent component C2 (humic-like DOM).

To explore whether there are thresholds along the hydrologic continuum that modulate the relative importance of hydrology and environment sorting, we analyzed the changes in NMDS1 scores as a function of WRT (Fig. 1.6b), and found that above a 10-day threshold, an increase in WRT did not trigger further changes in community composition along the NMDS1 axis. Thus, in order to compare the relative influence of the environmental and hydrologic conditions on the dissimilarity between communities within the NMDS ordination space, we performed a variation partitioning analysis dividing the dataset into two groups of sites based on whether their estimated WRT was >10 or <10 days (Fig 1.6c). This exercise clearly shows that environmental variables explained most of the variation (65%) in bacterial taxonomic composition at those sites

with high WRT (>10 days), whereas hydrological variables explained a larger share of the variation (33%) than the environment (14%) in sites with shorter WRT (Fig. 1.6c).

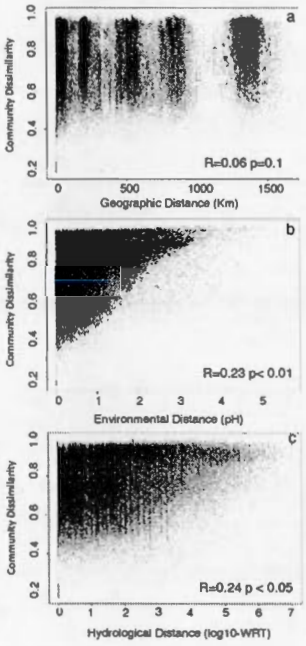


Figure 1.4. Relationship between geographical (a), environmental (b), and hydrological (c) distances and pairwise bacterial community dissimilarity (Bray-Curtis). pH and water residence time (log10-WRT) were used as proxies of environmental and hydrological distances, respectively. Mantel correlations (R) an the probabilities are provided for each case.

1.5 Discussion

Our results show that environmental sorting is a major driver of bacterial community assembly across the boreal biome, but further indicate that hydrology and network position, and the associated mass effects, also play a fundamental role in shaping bacterioplankton communities in complex aquatic networks. Although previous studies had identified these two processes as major mechanisms of freshwater bacterial community assembly (Crump et al., 2007; Fierer et al., 2007; Nelson et al., 2009; Ren et al. 2015, Souffreau et al. 2015), no study to date had explored how they interact to shape the large-scale biogeography of freshwater bacterioplankton communities. Assessing this interaction necessarily requires expanding the conceptual framework proposed by Martiny et al. (2006) to incorporate, in addition to the environmental gradients and geographic distance (as a proxy for dispersal limitation), a hydrologic dimension that controls the relative importance of the mass effects on community assembly, as we conceptualize in Figure 1.1.

Our results show no significant relationship between geographic distance and the taxonomic dissimilarity among the studied bacterioplankton communities (Fig. 1.4a), suggesting that dispersal limitation between aquatic bacterial communities is negligible across boreal Québec, at least at the taxonomic resolution of our analysis. Rather, there was an overall positive relationship between environmental and taxonomic distances (Fig. 1.4b), and communities appeared to be strongly structured by local water chemistry (mostly pH, Fig. 1.3b), pointing to a major role of environmental sorting in structuring the large-scale biogeography of boreal freshwater bacterioplankton. We should note that although pH emerged as the most important environmental variable, in agreement with previous studies (Fierer et al., 2007; Ren et al., 2015, Ruiz-González et al. 2015b), it is likely that besides directly influencing cell physiology, pH is also an integrator of other variables of relevance for bacteria (Fierer et al. 2007). In particular, we found that changes in water pH were strongly correlated to shifts in features linked to the amount and quality of DOM (details not shown), also known to shape bacterial community structure and functioning (Fujii et al., 2012; Ruiz-González et al., 2015a; Wilhelm et al., 2015).

Despite this strong effect of pH, the fact that in environmentally similar sites we observed the entire range of community dissimilarity (Fig. 1.4b) suggests that mechanisms other than environmental sorting are involved in shaping the large-scale biogeography of boreal freshwater bacterioplankton communities. In this regard, our results demonstrate that water residence time alone explains as much variability in composition as pH (Fig. 1.4b, c), and that bacterial communities change largely along a hydrologic continuum from headwater streams towards large rivers and lakes, following variations in water residence time and network position (Fig. 1.3, 1.5).

We hypothesize that this directional structuring is caused by a gradient of decreasing mass effects and increasing environmental sorting towards downstream ecosystems as conceptualized in Figure 1.1. In a companion study where we characterized the soil communities surrounding aquatic systems within a single boreal region (Ruiz-González et al., 2015b), we show that whereas headwater streams are strongly dominated by bacteria washed from soils, which gradually disappear along the hydrologic continuum, downstream ecosystems are numerically dominated by a few of these soil-derived taxa that have the potential to grow in the aquatic environment.

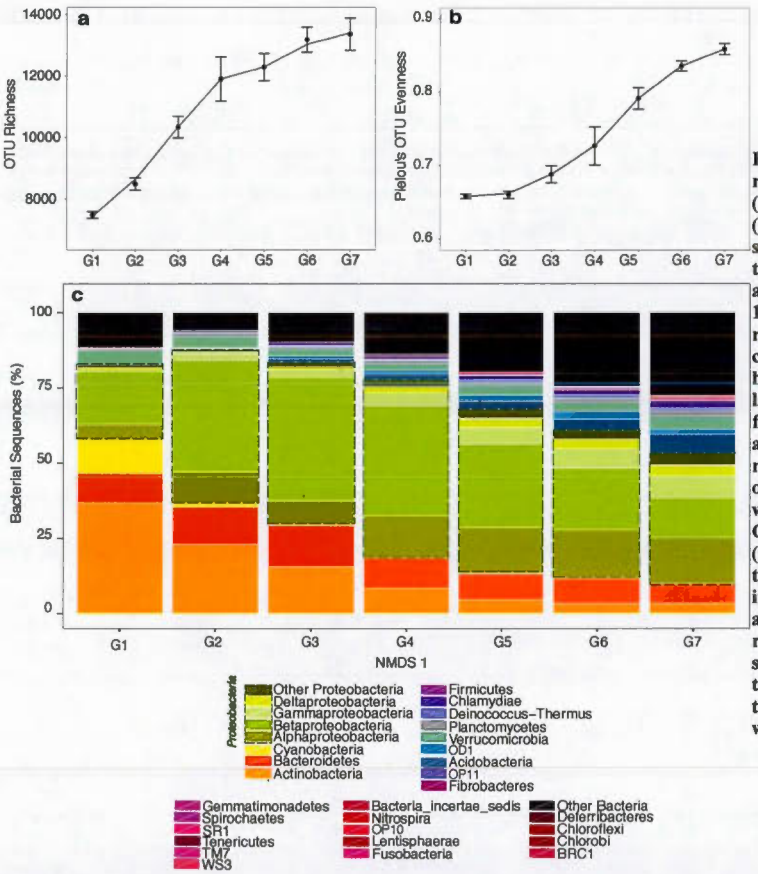


Figure 1.5. Changes in OTU richness (a), Pielou's evenness (b) and taxonomic composition (at the Phylum level, C) for sites grouped according to their position on the NMDS1 axis (G1 to G7, n = 174, 74, 37, 18, 26, 39, 18), which mostly represents an hydrologic continuum from the smallest headwater streams (G7) to larger rivers and lakes (G1, for details see Results). Dots are means and error bars represent the standard error of the values for the sites within each NMDS1 group. Colors indicate different Phyla (and different Classes within the Phylum Proteobacteria, indicated by the dashed line) and the heights of the bars represent the percentage of sequences associated to each taxonomic rank relative to the total number of sequences within each group of sites.

The high OTU richness found in our headwater streams across the boreal biome (Fig. 1.5a) thus likely results from the massive immigration of terrestrial bacteria. Moreover, we found that variations in d-excess (our proxy of groundwater-mediated connectivity with land) also explained changes in community composition (Fig. 1.3b), supporting that these mass effects depend on the degree of connectivity to the source of immigrants (i.e, the soils). In turn, the observed loss of bacterial taxa towards downstream ecosystems was accompanied (and probably caused) by gradual increases in the relative abundances of some common freshwater bacterial groups (sensu Newton et al. 2011, see examples in Supplementary Fig. S1.1), suggesting that environmental sorting progressively increases along the hydrologic continuum. In support of this idea, and in accordance to our conceptual framework (Fig. 1.1), we observed that the spatial patterns of the bacterial assemblages inhabiting systems

with longer WRT were much more strongly influenced by differences in local aquatic conditions than those from communities inhabiting headwater streams (Fig. 1.6a).

Interestingly, the relative influence of hydrology and the associated mass effects on community composition was not continuous: we observed a clear threshold of water retention time (WRT) of around 10 days, beyond which increases in WRT did not result in further hydrologically driven changes in taxonomic composition (Fig. 1.6b). Our variance partitioning analysis supports this hypothesis, demonstrating that hydrology explains a much larger fraction of the variance in taxonomic composition in systems below this WRT threshold, whereas environmental factors were responsible for most of the variability in ecosystems with longer WRT (Fig. 1.6c). In the latter systems, pH seemed to strongly influence OTU richness, which peaked at pH levels close to neutrality (Supplementary Fig. S1.2a), a pattern that has been observed before (Fierer and Jackson, 2006; (Fierer & Jackson, 2006; Ren et al., 2015). This suggests that extreme environmental conditions may impose a stronger filter for bacteria, and that neutral pH might favor the co-existence of taxa with overlapping ranges of tolerance. This relationship was lost, however, at sites with WRT>10 days (Supplementary Fig. S1.2b), indicating that intense mass effects may overwhelm species sorting even under extreme environmental conditions. Remarkably, we identified a similar WRT within a regional aquatic network (Ruiz-González et al., 2015b), beyond which there was no further selection or growth of soil-derived bacteria along the hydrologic continuum. This threshold is likely determined by the balance between the potential for in situ growth of the resident bacterial communities relative to the rate of dilution by allochthonous bacteria (e.g., Crump et al. 2007), but this issue needs further investigation.

Considering that the headwater systems are largely subjected to mass effects, we argue that the bacterial composition in these communities must reflect the nature of the terrestrial source communities that feed these networks. The question then becomes whether there are distinct pools of immigrant bacteria separated by dispersal barriers, or whether all aquatic communities derive from a common pool of taxa, at least at the scale of our study. The lack of a relationship between community dissimilarity and geographic distance discussed above suggests that the boreal biome represents a single bacterial province sensu Martiny et al. (2006), and thus that that there is low dispersal limitation

of taxa between regions. We are aware, however, that there could be dispersal limitation at finer taxonomic resolution, although in a previous study we have shown that the level of taxonomic aggregation used here (97%) resulted in the clearest spatial and temporal patterns in community composition from boreal and temperate lakes (Cottrell et al., 2015).

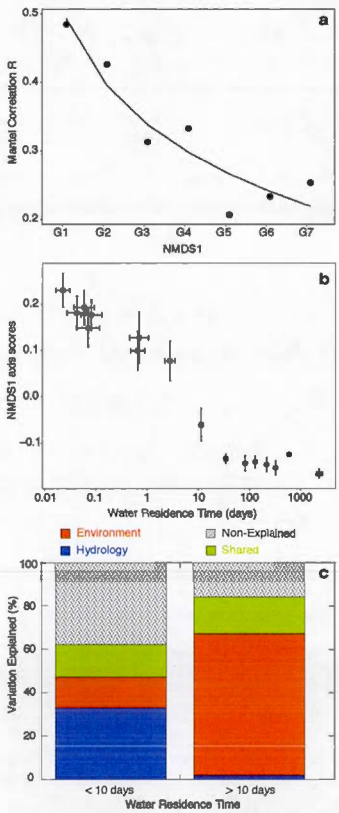


Figure 1.6. (a) Variation in the R coefficient of the Mantel correlations between the taxonomic and the environmental (pH) dissimilarity matrices for sites grouped according to their position on the NMDS1 axis (G1 to G7). Note that G1 includes most lakes and the largest rivers, and G7 (G1 to G7, n = 174, 74, 37, 18, 26, 39, 18) contains mostly small headwater streams (for details see Results) All the correlations were significant (p <0.01). (b) Changes in bacterial community composition along the NMDS1 axis as a function of water residence time (WRT). The NMDS1 scores of the sites were binned into 16 equal groups (n=20) based on ranked WRT. The dots are the average NMDS1 scores within each bin and represent the mean position of the sites within a given bin on the NMDS1 axis. Error bars are the standard errors for either WRT and scores within each bin. Note that increases in WRT result in gradual changes in taxonomic composition along the NMDS1 axis until a WRT of 10 days is reached, above which increases in WRT do not translate into further changes in taxonomic composition along the NMDS1 axis. (c) Percentages of variation in the NMDS ordination space explained by the environment and hydrology, for sites with water residence times above and below 10 days (WRT>10 and WRT<10, respectively). Non-explained and shared variation (environment + hydrology) are also shown. The "environment" category included the variables [DIC] dissolved inorganic carbon; [cDOM] colored dissolved organic matter; [%C2] percentage of fluorescent component C2 (humic-like DOM); [%C5] percentage of fluorescent component C5 (freshly produced labile DOM) and conductivity. [T] Water temperature; [CA] Catchment area; [AWC] area of water in the catchment; [WRT] Water residence time, and d-excess were included within the "hydrology" category.

In our conceptual framework, a common bacterial pool across all regions should lead to small differences in community structure between systems that are overwhelmingly dominated by mass effects (Fig. 1.1b), a scenario that is not supported by our results. Rather, we observed that the taxonomic differences between systems with the lowest WRT were as large as those between lake communities (Fig. 1.3), a pattern that resembles that of Figure 1.1a, where dispersal limitation plays a major role. Since our data suggest that this is not the case, we hypothesize that the observed differences in taxonomic composition between small headwater streams reflect a sorting that occurs

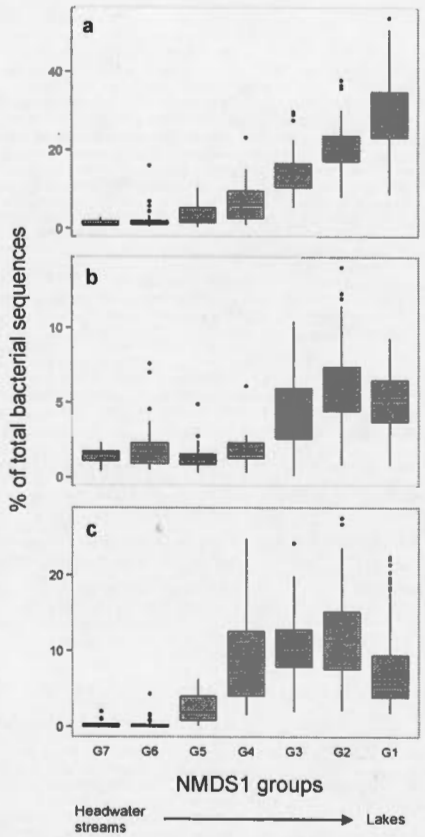
in the soils prior to the loading of terrestrial bacteria to the aquatic network. This may indicate the occurrence of a shared pool of taxa that undergoes re-arrangements driven by local landscape conditions. Thus, although there is evidence that even in headwater streams there is already some degree of aquatic sorting of terrestrially-derived taxa (Ruiz-González et al. 2015b), we argue that the communities from headwater aquatic systems that are subjected to strong mass effects must largely reflect differences in the composition of the surrounding soil assemblages. In this regard, the regional signature observed in headwater streams (Fig. 1.3a) could easily be interpreted as evidence of dispersal limitation (Fig. 1.1a). However, the fact that low-pH headwater streams from Schefferville clustered together with equally acidic sites in Côte-Nord (Fig. 1.3a,b), for example, suggests that the observed patterns are the result of a regional structuring of pH (and likely other covarying factors), and therefore, of pH-driven differences in soil bacterial taxonomic composition (Fierer and Jackson, 2006). This could be considered analogous to historical or legacy effects (Martiny et al., 2006) because it occurs outside the aquatic network and prior to the aquatic sorting, but is nevertheless based on contemporary environmental ranges and not necessarily on past events linked to dispersal limitation. Once in the aquatic network, these migrant communities undergo further selection by local conditions as they move downstream, and especially in systems with long water residence times (Crump et al., 2012; Ruiz-González et al. 2015b), such that the composition of these communities will be mainly a reflection of the local aquatic environment (e.g. Yanarell and Triplet, 2005, Fujii et al., 2012; Ren et al., 2015; Souffreau et al. 2015). This process renders lake or large river communities clearly distinct from those of headwater streams (Fig. 1.3, 1.5) within a given region, even when they may share similar ranges in key environmental factors. We recognize that our results provide a midsummer snap-shot and do not incorporate the potential temporal variability within communities, but when the taxonomic composition was analyzed for a subset of rivers and lakes sampled on 3 occasions (spring, summer, and fall, n=331, data not shown), the same pattern was observed: irrespective of the season, aquatic communities were structured along an hydrologic and a pH gradient, highlighting the consistency and strength of the spatial patterns presented here.

Overall, we show that hydrology and network position interact with environmental gradients to shape the spatial patterns of bacterioplankton community composition

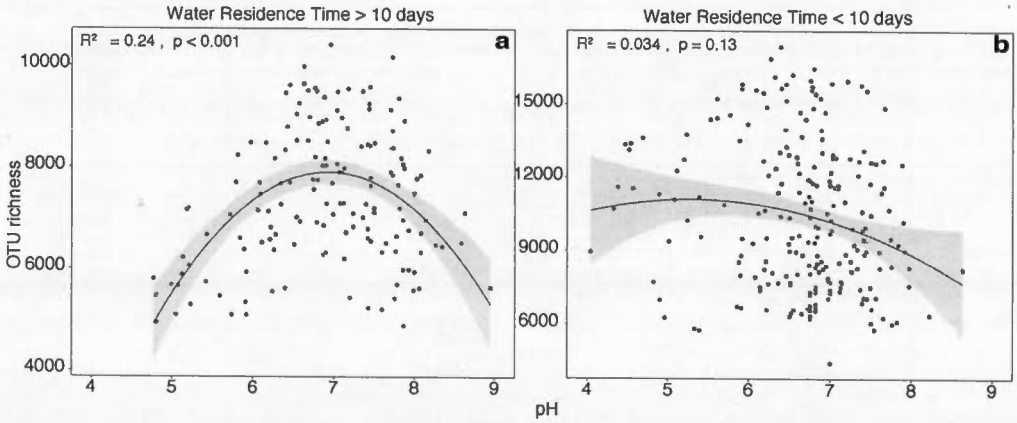
across boreal freshwater networks. Our results indicate that hydrology and network position modulate the relative role of environmental sorting and mass effects on community assembly along the aquatic continuum, likely by defining both the time frame for bacterial growth and the composition of the immigrant pool. We found no evidence of dispersal limitation even at the large spatial scale of our study, but we observed a certain regional signature likely caused by the spatial structuring of environmental conditions in the landscape. We suggest that this landscape structuring influences the headwater communities through a pre-sorting of the immigrant pool of bacteria, and the communities within systems with longer WRT by determining the local aquatic conditions that bacteria will face. Despite that these two processes can generate similar spatial patterns, they represent fundamentally different mechanisms, highlighting the fact that interpretations based on partial fragments of this complex biogeographic reality may lead to a biased understanding of the factors shaping bacterioplankton communities over large spatial scales.

1.6 Acknowledgements

We thank Ryan Hutchins for help with GIS analyses, Annick St-Pierre, Alice Parkes, and the whole CarBBAS team for their contribution to the field and laboratory components of this research, and Nicolas Fortin St-Gelais for critical comments and discussion. This study is part of the program of the Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBAS) Industrial Research Chair, co-funded by the Natural Science and Engineering Research Council of Canada (NSERC) and Hydro-Québec.



Supplementary Figure S1.1. Three examples of bacterial taxonomic groups that showed pronounced increases in their relative abundances along the hydrologic continuum: a) Order Actinomycetales (Cl Actinobacteria); b) Chitinophagaceae (Cl. Sphingobacteria); c) Genus Polinucleobacter (Cl. Betaproteobacteria). Values are calculated for the seven groups of sites according to their position on the NMDS axis (G1 to G7), which mostly represent a hydrologic continuum from the smallest headwater streams (G7) to larger rivers and lakes (G1, for details see results)



Suplementary Figure S1.2: Bacterioplankton operational taxonomic unit (OTU) richness along the pH gradient for sites with water residence times above (a) and below (b) 10 days (WRT>10 and WRT<10,respectively. R-squared and probabilities of the relationship are provided for each case.

CHAPTER II

LANDSCAPE-SCALE SPATIAL ABUNDANCE DISTRIBUTIONS DISCRIMINATE CORE FROM RANDOM COMPONENTS OF BOREAL LAKE BACTERIOPLANKTON

Juan Pablo Niño-García[1,2], Clara Ruiz-González[1], Paul A. del Giorgio[1]
Published in ECOLOGY LETTERS (2016), 1506 -1515.

1. Groupe de recherche interuniversitaire en limnologie, Département des Sciences Biologiques, Université du Québec à Montréal, Case Postale 8888, succursale Centre-Ville, Montréal, QC, H3C 3P8, Canada.

[2]Escuela de Microbiología, Universidad de Antioquia, Ciudad Universitaria Calle 67 No 53-108, Medellín, Colombia.

N.B References cited in this chapter are presented at the end of the thesis.

2.1 Abstract

Aquatic bacterial communities harbour thousands of coexisting taxa. To meet the challenge of discriminating between a "core" and a sporadically-occurring "random" component of these communities, we explored the spatial abundance distribution of individual bacterioplankton taxa across 198 boreal lakes and their associated fluvial networks (188 rivers). We found that all taxa could be grouped into four distinct categories based on model statistical distributions (normal-like, bimodal, logistic and lognormal). The distribution patterns across lakes and their associated river networks showed that lake communities are composed of a core of taxa whose distribution appears to be linked to in-lake environmental sorting (normal-like and bimodal categories), and a large fraction of mostly rare bacteria (94% of all taxa) whose presence appears to be largely random and linked to downstream transport in aquatic networks (logistic and lognormal categories). These rare taxa are thus likely to reflect species sorting at upstream locations, providing a perspective of the conditions prevailing in entire aquatic networks rather than only in lakes.

2.2 Introduction

A decade of genomic studies has revealed a recurrent pattern across natural bacterial communities, which appear to be composed of a few dominant phylotypes and an extremely large tail of rare taxa (Sogin et al., 2006). Even though this species dominance-abundance relationship represents one of the earliest and most general patterns recognized in ecology (McGill et al., 2007, Mathews & Whitaker, 2015), microbial communities are exceptional in the high proportion of rare taxa, which represent the largest reservoir of biological diversity in the biosphere (Pedrós-Alió, 2012; Locey & Lennon, 2016). Most studies on large-scale bacterial distributions to date have focused on the dominant taxa (Fierer & Jackson, 2006; Hanson, et al., 2007; Martiny et al., 2006; Nemergut et al., 2011; Newton, et al., 2011), but there is increasing interest in elucidating also the nature of the vast numbers of rare taxa in microbial communities. One challenge in meeting this goal is that the distinction between rare and dominant

components of the communities is not always clear. The reason is that even though some taxa persist as dominants in a given environment (Barberán, et al., 2011; Liu, et al, 2015; Nemergut et al., 2011) and others remain permanently rare, many appear to shift from rare to abundant and vice versa (Alonso-Sáez, et al, 2015; Alonso-Sáez et al., 2014; Baltar et al., 2015 Shade et al. 2014; Neueschwander et al. 2015).

It follows that rarity or dominance of taxa are not necessarily fixed characteristics within bacterial communities, and these variations are likely reflected also in the spatial distribution of bacteria. Most spatial studies of bacterial communities to date have focused on differences in bacterial communities among similar types of ecosystems, or on the variation of dominant taxa or specific bacterial groups along environmental gradients. These studies have revealed recurrent biogeographic patterns at different phylogenetic levels (Nemergut et al., 2011; Newton et al., 2011; Jones et al., 2012), taxon-specific differences in the environmental range of preferences (Evans & Wallenstein, 2013; Fierer, et al., 2007a; Lennon, et al., 2012; Newton et al., 2011), and clusters of taxa showing similar associations with environmental conditions (Barberán et al., 2011; Eiler, et al., 2011; Steele et al., 2011; Williams, et al., 2014). These approaches, however, have been insufficient to elucidate the patterns or significance of rarity in bacterial communities.

An alternative approach that has seldom been applied to microbial communities is the analysis of the spatial abundance distributions of individual species (SpADs) across sites or ecosystems (Harte et al., 2005; Conlisk et al., 2012). For macro-organisms, the typical shape of SpADs varies from strictly downward-sloping to increasingly normal-like, associated with enlarging the sampling count area (Conlisk et al., 2012). These different shapes must result from a combination of eco-physiological traits of taxa, which determine potential growth rates, environmental ranges of tolerance, and capacity to be dispersed and persist under unfavorable conditions. It may thus be possible not just to infer key aspects of the niche of individual taxa, such as modes of dispersal, but also to distinguish groups of taxa that share similar patterns of commonness and rarity (Conlisk et al., 2012), as has been shown for several animal communities (Magurran & Henderson, 2003; Ulrich and Ollik, 2004).

Applied to microbial communities, this approach would facilitate dissecting the bacterial rank abundance continuum into ecologically meaningful categories of taxa based on their patterns of spatial distribution. So far, only a few studies have modeled abundance distributions, mostly to identify temporal abundance patterns of aquatic bacteria (Shade et al., 2014; Shade & Gilbert, 2015), and to differentiate abundant and ubiquitous taxa from rare phylotypes (Galand et al., 2009). None, however, has attempted to characterize the structure of bacterial communities by modeling SpADs of the individual constituent taxa, which would allow distinguishing the environmental selection of bacteria from the passive dispersal-related processes that control local bacterial abundance. Discriminating between these two processes would in turn provide insight into the ecological underpinnings of rarity and commonness in bacterial communities. This discrimination is particularly challenging in freshwater ecosystems, where the presence of a taxon may not necessarily reflect growth but rather passive transport from upstream or adjacent terrestrial ecosystems (Crump, et.al., 2012; Niño-García, et al., 2016; Ruiz-González, et al., 2015b). In the case of lakes, identifying ecologically meaningful categories of bacteria with similar SpADs thus requires an approach that compares the abundance patterns of individual taxa across large enough environmental and geographic gradients, while placing these patterns in the context of the entire aquatic network, since the relative importance of dispersal-related (i.e. hydrologically driven transport) and niche-driven (i.e. species sorting by local conditions) processes can vary greatly across landscapes (Niño-García et al., 2016). Although until recently this kind of sampling design was technically impractical, the advent of high-throughput sequencing technologies at low cost now allows such large-scale comparisons across ecosystems at a meaningful sequencing depth.

Here we present the first comprehensive study of large-scale SpADs of lake bacterial communities. We analyzed the patterns in occurrence and abundance of individual taxa of lake bacterioplankton, characterized by Illumina sequencing of the 16S rRNA gene, across 198 boreal lakes of Québec that span large ranges in environmental conditions. We grouped these taxa into categories of similar spatial distribution across lakes, by fitting SpADs to a set of model statistical distributions. To assess the ecological underpinnings of the resulting categories, we explored the responses of taxa to the

environmental gradients covered by our study lakes, and retraced the patterns of distribution of each individual taxon within the fluvial networks associated with these lakes (188 additional river sites). This approach allowed us to distinguish between two main components of bacterial communities: A "core" composed of taxa whose SpADs suggest environmental sorting within lakes and a "random" component that comprises taxa whose SpADs suggest transport from upstream and persistence in downstream ecosystems despite a lack of or a low metabolic activity.

2.3 Materials and methods

2.3.1 Study sites, sampling design and basic parameters

The study area and sampling design have been previously described in Niño-García et al., (2016). Briefly, we collected water samples from 198 lakes and 188 rivers from seven regions of boreal Québec (Canada) that cover broad ranges in climate, hydrology and limnological properties (for details see Ruiz-González et al., 2015a; Niño-García et al., 2016), and a total area of 900,000 km^2 (44–56°N, 64–80°W). The sites were sampled once during summer (July/August) between 2009 and 2013, and span large ranges in lake area (0.002–4345 km^2) and river order (Strahler order 0–8). We defined as streams of order 0 those that did not appear in topographic maps and were likely temporary streams.

All sites within a region were sampled during the same year. Water samples were collected at a depth of 0.5m at the deepest measured point of lakes, and near the shore in rivers. Temperature, dissolved oxygen, pH and conductivity were recorded in situ with a multiparameter probe (600XLV2-M, Yellow Springs Instruments, Yellow Springs, OH, USA).

2.3.2 Bacterial community composition

DNA was extracted from 0.22-μm pore-size filters after filtering 300-500 ml of water using the MoBio PowerWater DNA extraction kit. We used 515F and 806R primers to amplify the variable regions V3-V4 of the 16S rRNA gene and sequenced them on an Illumina MiSeq2000 following a paired-end approach (Caporaso et al., 2012). Paired-end reads were assembled with FLASH (Magoč & Salzberg, 2011) and 250-290 bp sequences were used for downstream analyses in QIIME to remove primers and low quality, archaeal and chloroplast reads (Caporaso et al., 2010). We detected and removed chimeric sequences using the UCHIME algorithm (Edgar et al., 2011). To avoid using arbitrary thresholds of similarity, we clustered the sequences into OTUs using the SWARM 2.0 algorithm (Mahé et al., 2014), applying a local clustering threshold of d=1, and the fastidious option that reduces the number of singletons and doubletons by grafting them onto more abundant clusters (Mahé et al., 2015). We assigned taxonomy to the representative sequence of each swarm using the SILVA 111 reference database (Quast et al., 2012) and the RDP classifier (Wang, et al. , 2007). We discarded all OTUs with <10 sequences and/or present in <10 samples, and randomly subsampled the OTU table to generate an equal number of sequences per sample based on the sample with the fewest reads (41,112). Raw sequences have been deposited in the European Nucleotide Archive, under the accession number PRJEB14062.

2.3.3 Modeling bacterial SpADs

We used an iterative procedure that evaluates the statistical distribution that best fits the SpADs ($\log_{10}(x+1)$-transformed) of each OTU across the 198 lakes. We first identified all OTU abundance distributions that deviated from unimodality by computing the Hartigans' dip test statistic as implemented in the diptest package in R (Maechler, 2015); any OTU with p-value <0.01 was categorized as having a bimodal distribution. For all unimodal OTUs we fit several potential model distributions (normal, Weibull, logistic, gamma, lognormal, Cauchy and exponential) using maximum likelihood

estimation with the function distfit of the fitdistrplus package in R (Delignette-Muller & Dutang, 2014). This function also provides the Akaike Information Criterion (AIC) values that we used to select the model that best fit the observed abundance distribution of each OTU (i.e., the one with the lowest AIC). Most abundance distributions included frequencies of zero abundance, and thus the range of statistical models tested were explicitly selected to capture SpADs that included absences, ranging from symmetrical model distributions with various degrees of dispersion (normal, Weibull, Cauchy) to positively skewed model distribution with various degrees of asymmetry (lognormal, logistic, exponential).

To assess links between the taxonomic composition of the resulting categories of SpAD and environmental factors, we calculated the correlation between the compositional dissimilarity within each of the categories and environmental distance using Mantel tests (R Vegan Package, Okasanen et al., 2015). We further estimated spatial OTUs turnover within each of the spatial categories using the beta.pair function, as the turnover-fraction of Jaccard pair-wise dissimilarity (R betapart package; Baselga & Orme, 2012).

2.3.4 Ecological features associated with different SpADs

We calculated the ranges in relative abundance, occurrence, and environmental breadth of the OTUs belonging to different SpAD categories. In previous studies we found that pH is a key environmental variable differentiating bacterial community compositions in these boreal aquatic ecosystems (Ruiz-González et al., 2015b; Niño-García et al., 2016), so we calculated environmental breadth as the pH range at which a particular OTU was present, independently of its abundance. Environmental breadths calculated based on variables such as chlorophyll, nitrogen, phosphorus, total dissolved and colored organic carbon and temperature showed very similar patterns (details not shown).

To further explore the ecological underpinnings of these SpAD categories, we analyzed

the dynamics of their constituent OTUs in the context of the whole aquatic network (198 lakes plus 188 associated streams and rivers). We performed this analysis along a gradient of river order, which integrates landscape position and water residence time (Niño-García et al., 2016). As a reference, we also included all the OTUs that were present in rivers but not detected in any of the studied lakes ('unique to rivers' OTUs, n=75,695). The latter provides a network "control" for non-responsive OTUs that do not grow during transport along the network and either die or are greatly diluted or outcompeted during transit and do not reach lakes in notable numbers (Fig 2.3).

2.4. Results

2.4.1 Categories of bacterial SpADs in lakes

The studied lakes displayed broad ranges in environmental variables such as pH and concentrations of dissolved organic carbon and nutrients, as well as drainage area and water residence time (Table S2.1). We obtained 20,242,838 high-quality sequences that clustered into 198,230 OTUs across the 198 lakes. After rarefaction, we retained 15,830,632 sequences (118,717 OTUs, ca. 5,200 OTUs per lake on average), which mostly belonged to the phyla Proteobacteria (41% of all sequences from lakes), Actinobacteria (18%), Verrucomicrobia (7%), and Bacteroidetes (4%, details not shown). These data result in a classic rank-abundance curve that declines steeply and has a long tail of increasingly rare taxa (Fig. S2.1).

The SpADs of all OTUs detected across lakes could be effectively captured by one of four basic model distributions: 1) normal-like (i.e. normal, Weibull), 2) bimodal, 3) logistic and 4) lognormal (Fig. 2.1a-d). Other distributions tested, such as exponential and Cauchy, failed to produce significant fits to the observed abundance distributions. The normal-like category comprised OTUs that displayed a unimodal-symmetrical or slightly negatively skewed shape with a strong tendency to vary around a central value that was systematically >0.1% of the local abundance (Fig. 2.1a). The bimodal category included OTUs with two clear density peaks that varied in their position along the

abundance axis, and in most cases, but not all, one of the modes corresponded to zero (Fig. 2.1b). The logistic and lognormal categories, in contrast, described OTUs with a characteristic zero abundance mode, generated by many absences (Fig. 2.1c-d). These two categories of SpAD differed from each other in that, in general, OTUs showing a lognormal distribution tended to have a much steeper decline along the abundance axis, and longer right tails (Fig. 2.1c) compared to those in the logistic category (Fig. 2.1d). Both categories included OTUs that were at the limit of detection of our sequencing depth in most lakes.

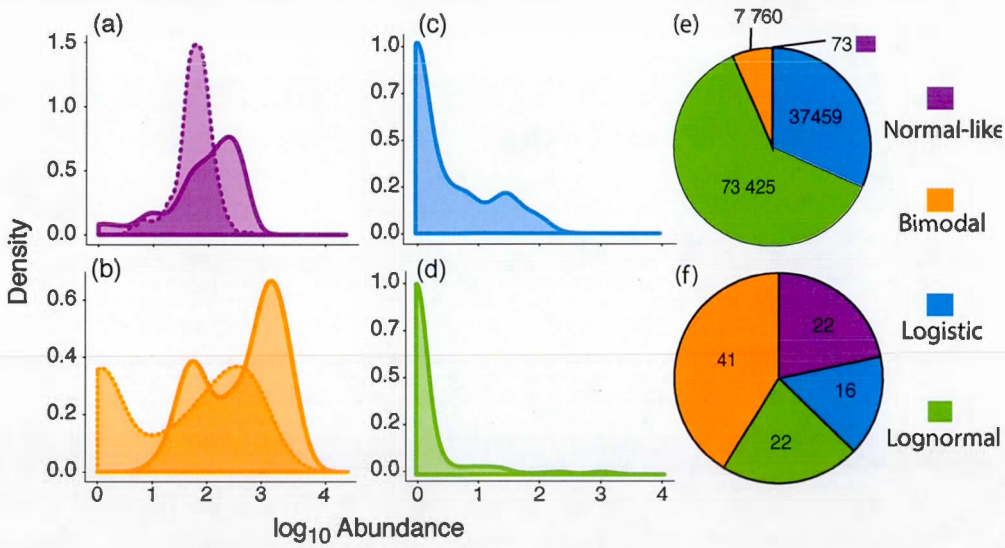


Figure 2.1. Examples of density plots of abundance distributions across 198 boreal lakes of representative OTUs belonging to different bacterial spatial distribution categories (a–d), and contribution of each SpAD category to the total number of lake OTUs (e) and associated sequences (f) in lakes. In (a) and (b) dotted and continuous lines indicate different individual OTUs within each category, and the height in the density plots represents the relative number of occurrences at a given class of abundance for each individual OTU

The lognormal and logistic categories comprised most (>90%) of the OTUs identified in lakes, yet in terms of abundance they accounted for only ca. 40% of the lake sequences (Figs. 2.1e–f). The normal-like and bimodal categories captured a small proportion of the total OTUs (0.6% and 7%, respectively) but together comprised more than 60% of the sequences found across all lakes (Figs. 2.1e,f). At a coarse level of

taxonomic resolution, there were no clear patterns in bacterial community composition between the four SpAD categories (Fig. S2.2). At finer taxonomic levels, however, we found clear differences, and some common freshwater groups such *Limnohabitans* and *Polynucleobacter*, for example, were much more prevalent in the normal-like and bimodal categories than in the other two (Fig. S2.3), and the dominant taxonomic groups clearly differed between the normal-like and bimodal categories (Table S2.2).

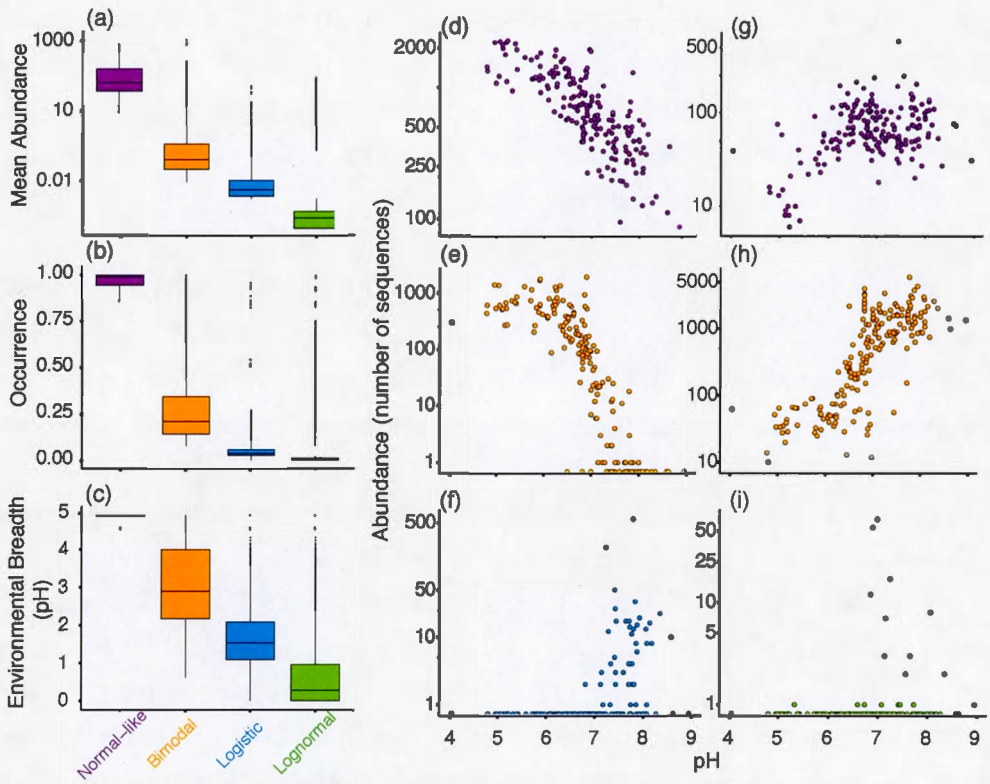


Figure 2.2. Mean abundance (a), occurrence (b) and environmental breadth assessed as pH range as a proxy (c) of OTUs within each SpAD category in lakes, and examples of abundance responses (i.e. changes in the number of sequences) of individual OTUs along the pH gradient in lakes (d–i). Different colours represent the four SpAD categories.

2.4.2 Ecological features of SpAD categories

Normal-like OTUs consistently exhibited the highest relative abundances, occurrences, and environmental breadth (Figs. 2.2a-c), whereas OTUs with logistic and lognormal SpADs were on average geographically more restricted, constrained to a narrower environmental range (i.e. pH range), and systematically found at very low abundances (Figs. 2.2a-c). OTUs within the bimodal category showed intermediate levels of abundance and highly variable occurrences (1-100% of sites), and also an intermediate environmental breadth relative to the other categories (Figs. 2.2a-c). None of the four categories was associated with a particular portion of the pH range: Within each spatial category, there were OTUs showing clear associations with acidic, circumneutral or basic conditions (see examples in Figs. 2.2d-h). This was also the case for the logistic and lognormal SpADs (Figs. 2.2f,i), although their low average occurrences renders the quantification of environmental breadth uncertain.

The analysis of distribution of the OTUs in each SpAD category within the whole aquatic network (i.e. 188 streams and rivers associated with the sampled lakes) showed that the average number of OTUs from lakes following a normal-like or bimodal distribution was constant or slightly increased, respectively, along the river continuum (Fig. 2.3a). However, these OTUs became major components in our study lakes, accounting for ca. 20% and 47% of all retained community sequences, respectively (Fig. 2.3b). In contrast, the mean contribution to all sequences in the lognormal and logistic categories steadily declined along the river continuum (Fig. 2.3b), although there was an offset between the two categories in lakes, where logistic OTUs were more abundant on average than lognormal OTUs, a pattern that was not observed in rivers (Fig. 2.4a,b). Both the number and proportion of lognormal OTUs originally present in headwater streams declined steeply along the river continuum, similar to the OTUs 'unique to rivers' (Fig. 2.3a and 2.4c, respectively). In contrast, the average number of OTUs per community belonging to the logistic category remained relatively constant along the river continuum (Fig. 2.3a), even when only those OTUs originally present in headwater streams are considered (Fig. 2.4c). This suggests some compensation of losses occurring along the river continuum and thus, that the differences between lognormal and logistic categories do not simply reflect differences in upstream

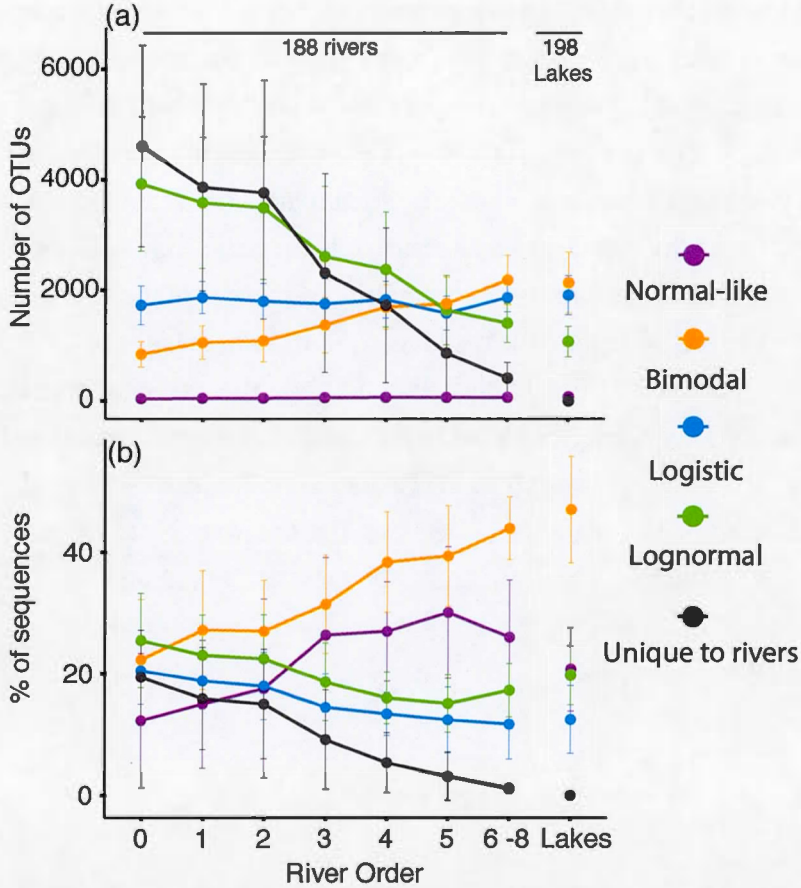


Figure 2.3. Comparison between the variation in the number of OTUs per community (a) and the percentage of community sequences (b) along the river continuum for each SpAD category, as well as the OTUs present in rivers but not detected in any lake (i.e. OTUs unique to rivers). Dots represent average values per type of ecosystem, and the error bars show standard deviations.

community composition.

2.4.3 Identifying distinct SpADs within the rare bacteria

The above results suggest that the typical SpAD of the rare components of lake bacterial communities is due to transport from upstream rather than to notable growth in lake environments. To explore whether there may be taxa that represent a class of constitutively rare lake bacteria, we extracted those OTUs that never surpassed

an arbitrary threshold of local relative abundance of 0.1%, regardless of the SpAD category to which they belonged, and recovered 116,799 OTUs. Although most of these OTUs had very low levels of occurrence, some showed occurrences as high as 90% (data not shown), suggesting that some widely distributed rare bacteria exist across our lakes. The 299 rare OTUs occurring at >60% of the sites (rare-but-ubiquitous category) differed greatly in their distribution along the aquatic network from the 116,143 OTUs occurring at <20% of the sites (rare-and-restricted category; Fig. 2.5). Both the OTU number and the average contribution to community sequences of the rare-and-restricted bacteria decreased steeply towards lakes (Fig. 2.5c-d), whereas those of the rare-but-ubiquitous taxa increased (Fig. 2.5a-b). In addition, the rare-and-restricted group mostly comprised OTUs belonging to the logistic and lognormal categories (Fig. 2.5c), whereas 99% of the rare-but-ubiquitous OTUs had bimodal SpADs (Fig. 2.5a).

Further examination of the correlation between the environmental distance matrices (using pH as an integrative proxy) and their respective taxonomic dissimilarity

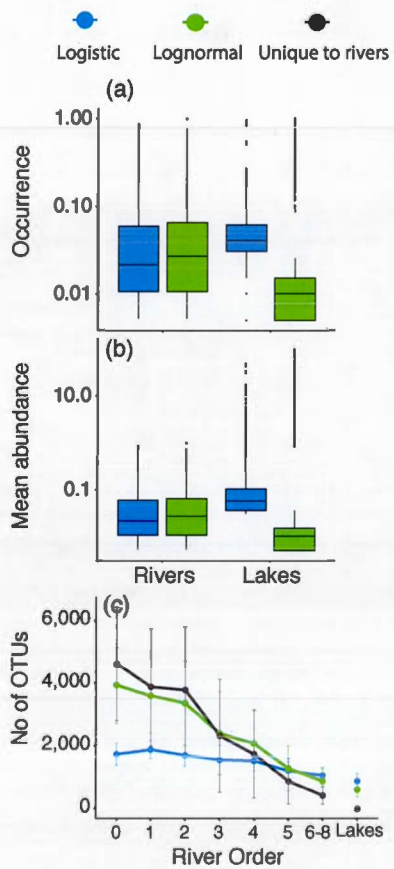


Figure 2.4. Comparison between the patterns of occurrence (a) and average number of sequences (mean abundance; b) for the logistic and lognormal SpAD categories, and changes in the number of OTUs in the logistic and lognormal categories from headwater streams to lakes (c) in comparison with the pattern shown by the OTUs unique to rivers. The analysis in panel c was restricted to OTUs detected in the smallest headwater streams to follow only the patterns of taxa present at the very beginning of the river continuum.

matrices, using a Mantel test, revealed associations of these two subgroups of rare taxa with environmental conditions. Since these associations could reflect community patterns upstream in the aquatic continuum or environmental sorting of bacterial taxa within lakes, this exercise was carried out separately for lakes and headwater streams (Strahler order <3). Although the correlation was stronger for the ubiquitous-rare than for the restricted-rare taxa (Mantel test, R=0.64 vs 0.49), both groups had significant Mantel correlations with pH in lakes. However, this correlation occurred also for the rare-and-restricted yet not for the rare-but-ubiquitous taxa in headwater streams (Fig. 2.6a). Moreover, the observed species turnover within the four categories across lakes (Fig. 6d) was lowest for the normal and bimodal categories (and within the last, the ubiquitous-rare group; Fig. 2.6c), and highest for the lognormal and logistic categories

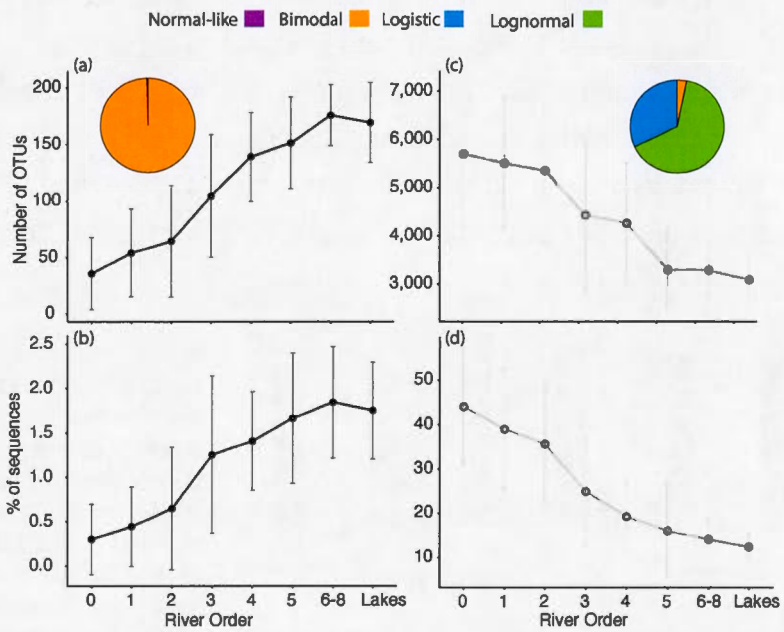


Figure 2.5. Changes along the river continuum in the number of OTUs and their contribution to sequences of rare-but-ubiquitous taxa (i.e. rare OTUs with occurrences > 60%) shown by black dots (a and b) and of rare-and-restricted OTUs with occurrences < 20% shown by grey dots (c and d). The pie charts in panels a and c show the percentage of OTUs within both rarity groups belonging to the different spatial categories (colours as in legend of Fig. 2.1).

as well as for the rare-and-restricted group (Fig. 2.6c-d).

2.5 Discussion

We could effectively categorize all lake OTUs into four discrete SpAD categories based on their abundance across 198 widely differing boreal lakes (Fig. 2.1). Our approach identified groups of taxa that show similar distribution patterns at the landscape scale, independent of their taxonomic affiliation or associations with particular environmental conditions. This allowed us to delineate core and random components of these communities. It is important to note that our approach differs from most spatial surveys in that it uses the spatial distributions of individual taxa to reveal patterns within the vast microbial diversity of natural ecosystems, rather than aiming to describe the composition of communities within or across ecosystems (e.g. Barberán & Casamayor, 2010; Lozupone & Knight, 2007; Nemergut et al., 2011) or environmental preferences of specific taxa (e.g., (Evans & Wallenstein, 2013; Fierer, et al., 2007a; Fierer, et al., 2007b). Although some studies have modeled individual abundance distributions of bacteria to identify specific temporal patterns (Shade et al., 2014; Shade & Gilbert, 2015) and to distinguish spatially abundant from rare taxa (Galand et al., 2009), ours is the first attempt to categorize and cluster bacterial OTUs based on their observed

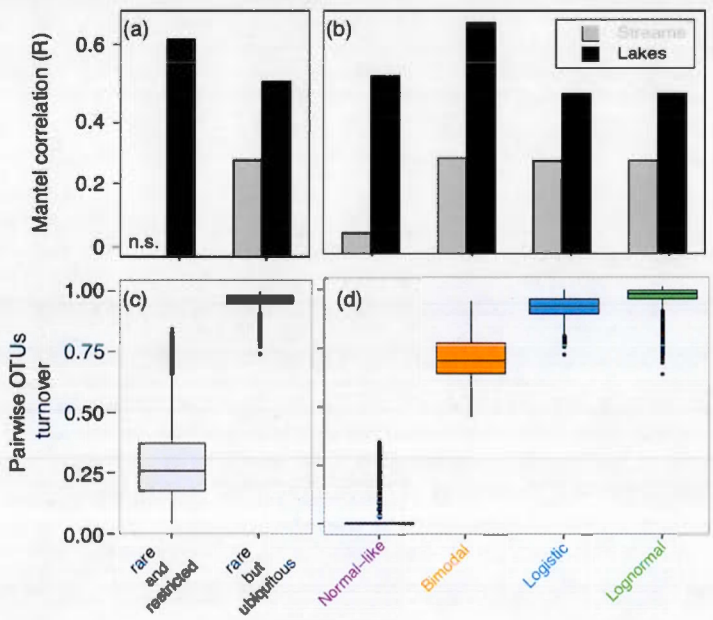


Figure 2.6. Mantel correlation between the environmental distance (pH) and taxonomic dissimilarity matrices for the two categories of rarity (a) and the four categories of SpAD (c) in small headwater streams (Strahler order <3, grey bars) and lakes (black bars). Fraction of pairwise dissimilarity between lakes due to OTU turnover for the two categories of rarity (b) and the four categories of SpAD (d).

SpADs without making a priori assumptions about abundance distributions.
Numerous terms have been proposed to denote different lifestyles associated with commonness and rareness, such as "natives" (Newton et al., 2011; Read et al., 2015), "tourists" (Zillio & Condit, 2007; Newton et al., 2011), "vagabonds" (Newton et al., 2011; Livermore et al., 2014; Read et al., 2015), "transients" (Magurran & Henderson, 2003; Ulrich & Ollik, 2004), "passive-streamlined" organisms and "responders" (Livermore et al. 2014). These terms are not always based on the same criteria and have different meanings for different authors. Therefore, we define here the terms "core" to denote taxa whose SpADs suggest environmental sorting within lakes, and "random" to refer to taxa whose SpADs suggest transport from upstream. These terms relate to the causes of occurrence rather than the degree of rarity, because some "random" taxa may attain significant abundances, whereas others that are consistently rare can form part of the "core" community in lakes.

Since lake bacterial communities can be markedly influenced by immigration from adjacent terrestrial or upstream aquatic ecosystems (Nelson et al., 2009; Crump et al. 2007;2012, Ruiz-González et al., 2015, Niño-García et al., 2016), the presence of a taxon in a lake does not necessarily imply preference for the prevailing environmental conditions. Consequently, to assign ecological meaning to these SpADs, we had to place them in the context of their distribution within the entire aquatic network, so that we could distinguish between SpADs that emerge because of tolerance of particular environmental conditions, resource utilization and growth within lakes, and those resulting from passive transport with downstream water flow. The fact that the OTUs within the four SpAD categories (Fig. 2.1) showed distinct ranges of mean abundance, occurrence and environmental breadth in lakes (Fig. 2.2a-c), together with their contrasting patterns along the river continuum (Fig. 2.3), support the notion that these SpADs result from different processes. For example, although most lake OTUs were already present in the network (only 2% of all lake sequences were exclusively found in lakes), normal-like and bimodal OTUs clearly increased in abundance towards lakes, whereas the number of OTUs and sequences of logistic and lognormal categories declined along the river continuum (Fig. 2.3). This is in accordance with previous observations that streams tend to have much higher bacterial richness than lakes, probably because of their tighter linkages with terrestrial environments, but that a large

fraction of these taxa is lost during transit along the river continuum and only a few can grow sufficiently to become major components in downstream lakes (Ruiz-González et al., 2015b; Niño-García et al., 2016). Although we have no data on growth or activity of these taxa, their individual SpADs suggest a first discrimination between OTUs whose distributions reflect recruitment, growth and environmental sorting within lakes (normal-like and bimodal categories) and those whose SpADs mostly reflect transport (lognormal and logistic categories).

It is interesting to note that most of the sequences within the normal-like and bimodal categories were associated with common freshwater bacterial groups (sensu Newton et al., 2011; Table S2.2), such as *Polynucleobacter*, *Limnohabitans* or the hgcI clade, which responded by shifting their relative abundances (Fig. 2.2d,g,e,h) and richness (Fig. S2.4) along the pH gradient in lakes. On the other hand, these freshwater groups were less represented in lognormal and logistic categories (Fig. S2.3), whose responses to pH gradients were smaller or negligible (Fig. 2.2f, i, Fig. S2.4) and whose shifts in composition seem to reflect mostly turnover of OTUs (Fig. 2.6d). All the evidence above supports the interpretation that these two core categories are shaped by local environmental sorting, and yet appear to represent two distinct ecological categories of bacterioplankton: The normal-like category comprises a relatively small number of OTUs (range 55-73 OTUs per lake) of highly abundant ubiquitous bacteria belonging to only 4 bacterial classes (Table S2.2), whereas the bimodal category includes many more groups (average of 2134 OTUs per lake assigned to 36 bacterial classes; Table S2.2) with more variable occurrences, associations with a narrower range of environmental conditions, and a much greater range in abundance (Fig. 2.2), including taxa that are always rare (Fig. 2.5a). Although the two types of bacteria have previously been shown to occur in fresh waters (Newton et al., 2011; Neuenschwander et al., 2014), our results further suggest that their relative contributions to the core bacterioplankton communities are remarkably consistent across lakes (Fig. 2.3b).

In contrast, the logistic and lognormal categories, which together accounted for >94% of the total OTUs detected in all lakes, comprise bacteria whose SpADs appear to be mostly due to persistence during downstream transport rather than in-lake taxa sorting. Although the key role of dispersal in structuring freshwater bacterioplankton

communities has been repeatedly shown (Crump et al., 2004; Lindström & Bergström, 2004; Lindström & Langenheder, 2012; Nelson et al., 2009), the almost complete species turnover observed in our study across lakes in the lognormal and logistic categories further supports the hypothesis that these OTUs in lakes occur largely sporadically because of downstream transport. However, unlike OTUs unique to rivers (Fig. 2.3), a significant fraction of the logistic and lognormal OTUs were still detected in lakes, suggesting that the SpADs of these taxa reflect persistence in lake environments. The contrasting patterns between these two categories within the aquatic networks that we investigated indicates additional differences in survival capacities. For example, the steep decline along the river continuum in the number and abundance of lognormal OTUs (Fig. 2.3, 2.4c) implies not only absence of growth, but also significant loss during transit. In contrast, logistic OTUs tend to persist, perhaps because of dormancy and particle association or other forms of protection (Jones & Lennon, 2010; Lutz, et al., 2013; Shade et al., 2012). Without this network context, these last two categories, which capture most rare taxa, might be considered as constitutively rare lake bacteria, but they are not. We acknowledge, however, that some of these taxa may become active at different times of the year (Lynch & Neufeld, 2015; Shade et al., 2013), and their logistic or lognormal SpADs might fit better to a bimodal or multimodal distribution over time. Therefore, the extent to which these SpADs are maintained over time deserves further investigation.

Not all rare bacteria followed the above patterns. We identified a small subgroup of rare-but-ubiquitous lake bacteria which were widespread but consistently rare across lakes and appeared to be filtered from the community during downstream transport in the river networks. This suggests that within the rare bacteria there may be taxa for which rareness is not random, perhaps related to slow growth or refuge and defense strategies (e.g. avoidance of predation or viral infection), as previously suggested (Logares, et al., 2015; Pedrós-Alió, 2012). This group of ubiquitous-rare bacteria contains very few taxa, all belonging to the bimodal category and with significant Mantel correlations with pH in lakes (Fig 2.6A), relative to the thousands of random OTUs that populate the lognormal and logistic categories (Fig. 2.5a, c), indicating that this form of bacterial rareness is an exception in lake communities. On the other hand, the high species turnover of rare-and-restricted bacteria (Fig. 2.6A) suggests that

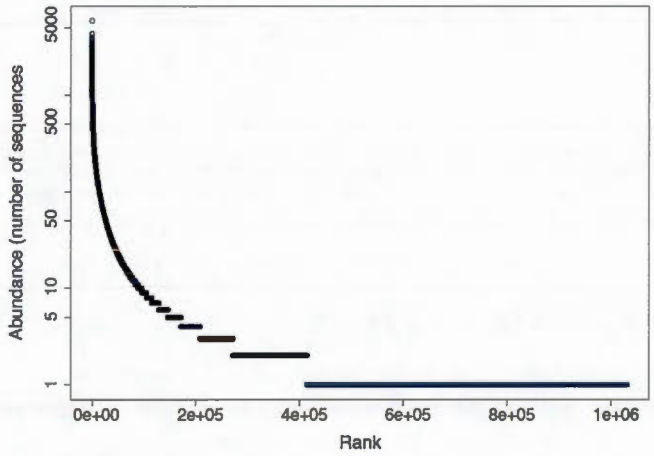
the observed relationship between the composition of this group and environmental conditions (Fig. 2.6c) is not necessarily due to environmental sorting of taxa within lakes, but to downstream transport in the aquatic network. The last is coherent with similar results found for SpAD categories that show higher species turnover and larger differences in the strength of Mantel correlations between streams and lakes for the core than for the random categories of lake bacterioplankton (Fig. 2.6c, d).

In summary, our results agree with analyses of non-microbial communities, which show a random component of community members related to dispersal and immigration, and a non-random core component related to species sorting, both of which may be inferred from the spatial distribution patterns of the taxa involved (Magurran & Henderson, 2003; Ulrich & Ollik, 2004). Moreover, we found that most rare taxa were grouped in the lognormal and logistic SpAD categories, which both reflect downstream transport but possibly differ in their patterns of dispersal, persistence and survival along the network. Although these rare taxa are present in lakes, they appear to result from species sorting elsewhere at upstream locations, thus providing an integrated perspective on the conditions prevailing in the entire aquatic network rather than in the lakes. Because network properties such as spatial configuration, retention, and terrestrial inputs vary across landscape types, the composition of the bacterial communities that transit these networks should vary as well. This may explain why the rare and apparently random components of these lake communities still exhibit some degree of spatial structuring and apparent environmental preferences (Fig. 2.6). Previous studies have reported biogeographical patterns even of rare microbes (Galand, et al., 2009; Liu et al., 2015; Logares et al., 2013; Vergin, et al., 2013), but the processes determining these patterns

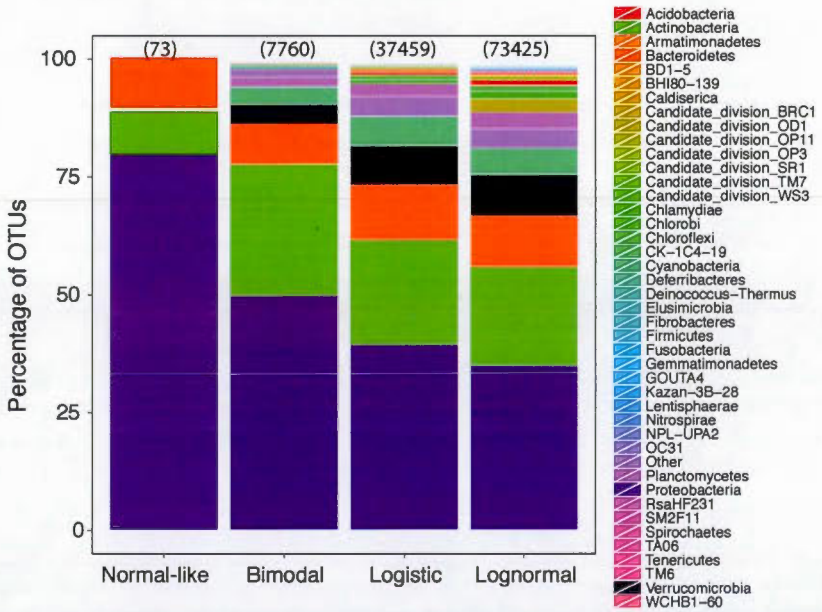
may be fundamentally different from those structuring the core component of these bacterial communities.

2.6 Acknowledgements

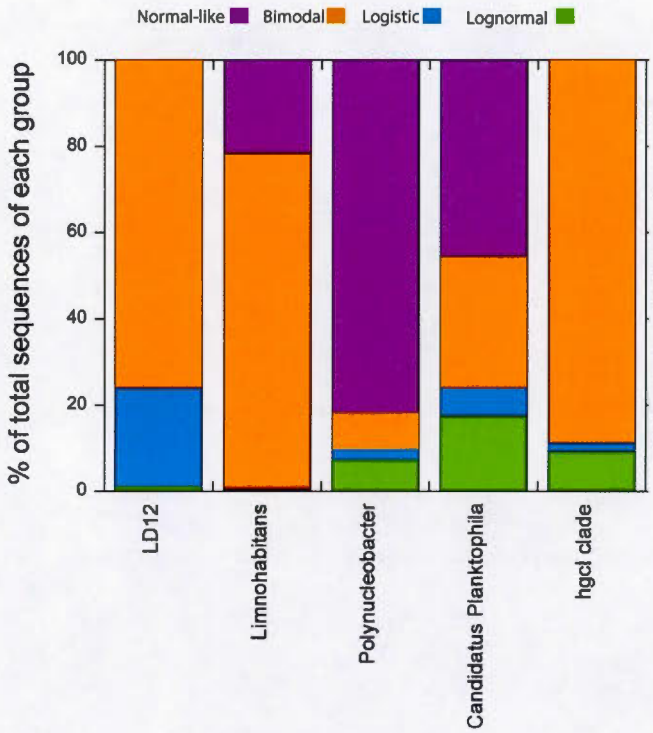
We thank Annick St-Pierre, Alice Parkes and the whole CarBBAS team for their contributions to the field and laboratory components of this research, and Richard Vogt for critical comments. We also thank the anonymous Reviewers and the Editors whose insightful comments improved this manuscript. This study is part of the program of the Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBAS) Industrial Research Chair, co-funded by the Natural Science and Engineering Research Council of Canada (NSERC) and Hydro-Quebec.



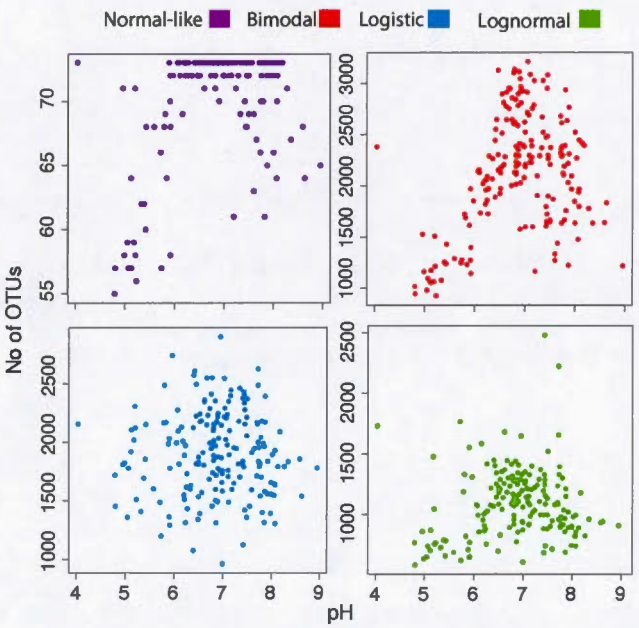
Supplementary Figure S2.1. Rank abundance curve



Supplementary Figure S2.2. Bacterial composition for each category of SpAD in 198 lakes. Each color is a different Phyla and the numbers above each bar are the OTUs included in the analysis for each category. Note that the major Phylum (i.e. Proteobacteria, Actinobacteria and Bacteroidetes) were represented in all the groups of SpAD and that they diversity are a function of the number of OTUs within each category. We observed the same pattern for Class and Order taxonomic levels (details not shown).



Supplementary Figure S2.3. Distribution of groups of SpAD categories (different colors) within five freshwater reprentative groups (LD12, Limnohabitans, Polynucleobacter, Candidatus Planktophila, hgcI clade) in 198 lakes. Note that these freshwater groups were dominated by normal-like and bimodal categories.



Supplementary Figure S2.4. Change in OTU richness across 198 lakes with different pH for different categories of SpAD (colors). Note that richness tend to peak in intermediate pH levels in normal-like and bimodal categories, suggesting that these categories are responding to the local environmental sorting.

Variable	Minimum	Maximum
pH	4.04	8.95
DOC (mg/L)	1.064	39.67
cDOM (m^{-1})	0.01	20.04
TN (mg/L)	0.1	0.91
TP (µg/L)	2.58	153
TEMP (C)	10.75	24.7
CHL (µg/L)	0.14	36.05
CA (Km^2)	0.002	4345
WRT(days)	1.38	10523
Zmax (m)	0.5	78
Secchi (m)	0.1	9.5

Table Supplementary 2.1. Range of main limnological and landscape variables measured in 198 lakes.

	Class	OTUs	Reads	%reads	Family	OTUs	Reads	%reads	Genus	OTUs	Reads	%reads
Normal-like												
	Betaproteobacteria	50	1218314	72.42635	Burkholderiaceae	14	639293	38.0	Polynucleobacter	14	639293	38.0
	Sphingobacteria	8	192131	11.42181	Comamonadaceae	25	402996	24.0	Sediminibacterium	8	192131	11.4
	Actinobacteria	7	189529	11.26712	Chitinophagaceae	8	192131	11.4	Candidatus Planktophila	3	144607	8.6
	Alphaproteobacteria	8	82168	4.884724	Sporichthyaceae	4	160877	9.6	LD28 freshwater group	4	135365	8.0
					Methylophilaceae	4	135365	8.0	Albidiferax	4	133198	7.9
					alphaI cluster	4	51201	3.0	Limnohabitans	6	47898	2.8
					Alcaligenaceae	4	28919	1.7	GKS98 freshwater group	4	28919	1.7
					Microbacteriaceae	3	28652	1.7	Candidatus Planktolibera	3	28652	1.7
					Sphingomonadaceae	1	15955	0.9	Sphingomonas	1	15955	0.9
					Oxalobacteraceae	3	11741	0.7	Piscinibacter	1	5429	0.3
					Unidentified families	3	15012	0.9	Unidentified Genera	25	310695	18.5
Bimodal												
	Actinobacteria	1993	1770749	46.4	Sporichthyaceae	1502	1604715	42.1	hgcI clade	694	1042824	27.4
	Betaproteobacteria	2264	814894	21.4	Comamonadaceae	789	420097	11.0	Limnohabitans	33	172112	4.5
	Alphaproteobacteria	1086	376727	9.9	Burkholderiaceae	405	161372	4.2	Candidatus Planktophila	185	96773	2.5
	Cyanobacteria	316	143781	3.8	Family1	307	139260	3.7	Limnobacter	27	96123	2.5
	Sphingobacteriia	501	139864	3.7	Chitinophagaceae	433	108105	2.8	Prochlorococcus	82	89764	2.4
	Spartobacteria	112	89208	2.3	FukuN18 freshwater_group	71	73675	1.9	Polynucleobacter	368	62746	1.6
	Cytophagia	88	68600	1.8	Acetobacteraceae	119	65960	1.7	Arcicella	46	61927	1.6
	Planctomycetacia	101	55052	1.4	Cytophagaceae	79	65912	1.7	Mycobacterium	47	52203	1.4
	Flavobacteria	100	50145	1.3	Microbacteriaceae	207	62669	1.6	Rhodovarius	43	50430	1.3
	Acidimicrobiia	123	43540	1.1	Planctomycetaceae	101	55052	1.4	Methylobacterium	80	49504	1.3
	Gammaproteobacteria	143	41455	1.1	Mycobacteriaceae	47	52203	1.4	Ferrigenibacter	46	47840	1.3
	Opitutae	118	28622	0.8	Methylobacteriaceae	91	50039	1.3	Flavobacterium	69	33270	0.9
	Deltaproteobacteria	89	24495	0.6	Sphingomonadaceae	87	37184	1.0	Candidatus Planktophila	66	21995	0.6
	Candidatus Methylacidiphilum	68	22888	0.6	Methylophilaceae	197	34970	0.9	CL500-29 marine group	40	21634	0.6
	OPB35 soil group	21	15401	0.4	LD12 freshwater group	34	33925	0.9	Sediminibacterium	181	21344	0.6
	Physcisphaerae	40	12219	0.3	Flavobacteriaceae	71	33393	0.9	Sphingomonas	31	19925	0.5
	Verrucomicrobiae	20	12049	0.3	Caulobacteraceae	102	32382	0.8	Novosphingobium	40	13471	0.4
	Deinococci	38	11672	0.3	Acidimicrobiaceae	45	22233	0.6	Opitutus	63	12908	0.3
	Chloroflexi	7	3756	0.2	Sinobacteraceae	42	17740	0.5	LD28 freshwater group	105	12241	0.3
	Thermoleophilia	27	5642	0.1	Alcaligenaceae	73	16350	0.4	CL500-3	38	12125	0.3
	Other classes	104	14623	0.4	Other families	940	261713	6.9	Other genera	880	272501	7.1
	Unidentified classes	401	65349	1.7	Unidentified families	2019	463773	12.2	Unidentified genera	4596	1549062	40.6

Table Supplementary 2.2. Taxonomic composition (at the Class, Family and Genus level) of OTUs within the normal-like and bimodal categories.'% reads' represents the relative contribution of each taxonomic group to the total number of reads within each category (i.e., normal-like or bimodal). For the case of the bimodal category, only the 20 most abundant groups are shown for simplification.

CHAPTER III

EXPLORING THE ECOLOGICAL COHERENCE BETWEEN THE SPATIAL AND TEMPORAL PATTERNS OF BACTERIOPLANKTON IN BOREAL LAKES

Juan Pablo Niño-García[1,2], Clara Ruiz-González[1], Paul A. del Giorgio[1]
Submitted to Frontiers in Microbiology

[1]Groupe de recherche interuniversitaire en limnologie, Département des Sciences Biologiques, Université du Québec à Montréal, Case Postale 8888, succursale Centre-Ville, Montréal, QC, H3C 3P8, Canada.

[2]Escuela de Microbiología, Universidad de Antioquia, Ciudad Universitaria Calle 67 No 53-108, Medellín, Colombia

N.B References cited in this chapter are presented at the end of the thesis.

3.1 Abstract

One of the major contemporary challenges in microbial ecology has been to discriminate between the reactive core from the random, unreactive components of bacterial communities. We had previously grouped lake bacterioplankton taxa into four distinct categories of spatial abundance distributions across boreal lakes of Québec, which appeared to reflect either hydrology-mediated dispersal along the aquatic network or environmental selection mechanisms within lakes. Here we test whether this categorization derived from the spatial distribution of taxa is maintained over time, by analyzing the temporal dynamics of the OTUs within these spatially-derived categories along an annual cycle in the oligotrophic lake Croche (Québec, Canada), and assessing the coherence of the ecological properties of OTUs within each spatial category over space and time. We found that the temporal dynamics of most taxa within a single lake are largely coherent with those derived from their spatial distribution over large spatial scales, suggesting that these properties must be intrinsic of particular taxa. Although we identified a few rare taxa catalogued as random from the spatial perspective but that showed clear seasonality and abundance peaks along the year, these comprised a very small fraction of the total rare OTUs. We conclude that the presence of most of the rare taxa in boreal lakes is random, since both their temporal and spatial dynamics suggest links to downstream transport and passive persistence in freshwater networks.

3.2 Introduction

It is now well established that most aquatic bacterial communities show a recurrent structure in which a relatively small number of abundant bacteria coexist with a vast number of extremely rare taxa that comprise the long tail of the so-called "rare biosphere" (Pedrós-Alió, 2006; 2012; Sogin et al., 2006). Whereas it is clear that the most abundant taxa within a given ecosystem tend to be reactive to local environmental conditions, be temporally and spatially dynamic, and play an active role in the community (Campbell et al., 2011; Fierer and Jackson, 2006; Fierer et al., 2007; Nemergut et al., 2011; Shade et al., 2013; Stuart E Jones, 2012; Van der Gucht et al., 2007), this becomes increasingly difficult to establish as we move deeper into the "rare biosphere", which presumably also includes taxa with different life strategies, adaptations, preferences, and sources (Campbell et al., 2011; Jones and Lennon, 2010; Lynch and Neufeld, 2015). For example, there may be rare bacteria that are part of a microbial "seed bank" and can activate to reach high abundances under specific environmental conditions (Aanderud et al., 2015; Alonso-Sáez, 2014; Alonso-Sáez et al., 2015; Baltar et al., 2015; Gibbons et al., 2013; Lennon and Jones, 2011; Wilhelm et al., 2015, Ruiz-González et al. in press), taxa that have intrinsically low growth and activity rates and for which rarity is a strategy (Campbell et al., 2011; Logares et al., 2015; Pedrós-Alió, 2012; Vergin et al., 2013), or taxa that are rare simply as a result of transport to the ecosystem, but which are not selected within aquatic ecosystems (Niño-García et al. in press, Hugoni et al., 2013; Vergin et al., 2013).

As the depth of sequencing (and thus the number of taxa that are detected in any given ecosystem) increases, discriminating between taxa that are selected and thus potentially active within the ecosystem, versus those whose presence reflects dispersal and random arrival becomes an ever more complex and pressing challenge. Most of the recent research addressing the issue of activity within aquatic microbial communities, and particularly within the rare biosphere, has focused on the analysis of ratios of 16S rRNA to rRNA genes (rDNA) of individual bacterial taxa, the application of probe-based cell analysis or stable-isotope probing within a limited amount of samples or in the exploration of temporal patterns of specific microbial populations (e.g. Aanderud et

al., 2015; Alonso-Sáez, 2014; Alonso-Sáez et al., 2015; Campbell et al., 2011; Hugoni et al., 2013; Vergin et al., 2013). These studies have provided fundamental insight into the distribution of activity and population dynamics within specific bacterial communities, but they do not provide a broad perspective on how reactive versus non-reactive taxa are distributed across aquatic ecosystems, particularly within the rare biosphere.

In a previous study (Niño-García et al. in press) we addressed this issue by using the abundance patterns of individual bacterial OTUs across a wide range of boreal lakes to group bacteria based on their individual spatial abundance distributions (SpADs). We found that all the OTUs (118,717) could be clearly clustered into four categories of SpADs, that we named after the model best explaining their abundance distribution (i.e., normal-like, bimodal, logistic and lognormal). Since freshwater bacterioplankton communities are known to be highly influenced by the immigration of taxa from upstream terrestrial or aquatic ecosystems (Crump et al., 2012; Niño-García et al., 2016; Ruiz-González et al., 2015; Savio et al., 2015), we explored the ecological underpinnings of the resulting categories by retracing these OTUs to the fluvial networks associated to these lakes, to distinguish between taxa whose presence and abundance is driven by downstream transport vs those subjected to environmental selection within lakes. Interestingly, two of these categories of SpAD (normal-like and bimodal) comprised the core of bacterioplankton communities, which included taxa whose patterns of spatial distribution were clearly linked to active growth within lakes, and to selection by local conditions. The other two categories (logistic and lognormal) grouped mostly rare bacteria whose presence was linked to downstream-mediated dispersal rather than in-lake selection and growth, and that were therefore categorized as "random" within lakes. Overall, we concluded that ca. 90% of OTUs, representing ca. 38% of all sequences found in boreal lakes, seem to be "random" (Niño-García et al. in press), and thus unlikely to be playing a significant role in the biogeochemical and trophic functioning of these ecosystems.

Since these categories were based on the SpADs of individual taxa across many lakes but that were measured on a single occasion (mid-summer), we do not know to what extent the ecological significance attributed to such categories is maintained over time. For example, it could be possible that the distribution of the OTUs that we found to be

ubiquitous across space and highly abundant was restricted to summer conditions, and thus on a temporal basis, they would be a summer occasional taxa (*sensu* Magurran and Henderson, 2003). Similarly, rare taxa classed as "random" based on their SpADs may in fact show active recruitment and abundance peaks at other times of the year (Shade and Gilbert, 2015; Shade et al., 2014). Should that be the case, it would imply that the spatial categorization does not capture the full spectrum of bacterial functional categories in lakes, and would require incorporating a temporal dimension to effectively discriminate between truly "random" versus core lake taxa. Conversely, if these SpADs are consistent at the temporal scale, the features that they reflect should be intrinsic properties of taxa.

Although studies with non-microbial organisms have shown that the rate of species accumulation in a given ecosystem results from the influence of both temporal and spatial components, suggesting that processes related with these components equally influence local community assembly and diversity scaling (Adler *et al.*, 2005; Adler and Lauenroth, 2003), only a handful of studies have explored the interaction between species abundances over time and space (Guo et al., 2000; Hanski and Gyllenberg, 1993; McGeoch & Gaston 2002). Interestingly, examination of th sparse evidence available suggests that infrequent species tend to have low abundances when they do occur whereas the most abundant are also more frequently found over time (Magurran, 2007), yet we do not know whether this is the case for microbial taxa. To our knowledge no microbial study has addressed whether specific strategies derived from spatial patterns of taxa are also maintained over time and vice versa. Thus, exploring the ecological spatial-temporal coherence of different categories of bacterial SpADs will provide information not only on how the temporal and spatial dimensions will influence the structuring of microbial communities but also about the mechanisms generating rarity and commonness.

In order to explore this issue, here we investigate the temporal dynamics of OTUs that had been previously grouped within four categories of SpADs (Niño-García *et al.* in press) along an annual cycle in oligotrophic lake Croche (Québec, Canada), and we assessed whether there is coherence between the properties of the OTUs within each category, over space and time. In order to extend these results beyond lake Croche, we

further compared the spatial and temporal patterns of taxa across 21 additional boreal lakes for which we had data for three seasons (spring, summer and fall). Overall, our result suggest that the ecological properties associated to the categories of SpAD based on large scale distribution patterns seem to be maintained over time and thus that they may be idiosyncratic for each bacterial taxon, at least in this boreal region.

3.3 Materials and methods

3.3.1 Study sites, sampling design, and bacterial community composition

In this study we present the results of a seasonal study carried out in lake Croche, located in the Laurentian region of Québec (47°24'N 71°47'W), where water samples were collected monthly from July 2012 to October 2013 for the characterization of bacterioplankton communities inhabiting the epilimnion. Lake Croche is a small (18.1 ha) and relatively deep (mean depth 6m) oligotrophic headwater lake surrounded by a pristine watershed. The lake thermally stratifies from late June through September, and is covered with ice from early October to May. The mean annual surface water temperature is 15°C, ranging from 2°C in January to 24 °C in July, with a mean water retention time is 1.1 year (Carignan et al., 2000; Vachon and del Giorgio, 2014).
In addition, during 2012 and 2013, we sampled 21 lakes from four different boreal regions in Quebec (Abitibi, Bay James, Chibougamau, Saguenay), each on three occasions (spring, summer and fall, n= 63). Although these lakes represent just a subset of the 198 lakes used in our previous spatial surveys (Niño-García *et al.* 2016; Niño-García *et al.* 2016), with this design we were able to capture reasonably well the variation in most limnological, environmental and hydrological characteristics (Supplementary table 3.1).

To characterize bacterial communities, DNA was extracted from 0.22 µm pore-size filters after filtering 300-500 ml of water using the MoBio PowerWater DNA extraction kit, following the manufacturer's specifications. We used the 515F and 806R primers to flank the variable regions V3-V4 of the 16S ribosomal RNA gene, which we sequenced

using Illumina MiSeq2000 using a paired-end approach (Caporaso et al., 2012). We then assembled the paired-end reads by using FLASH (Magoč and Salzberg, 2011) and we obtained sequences with 250-290 bp, that were used for the downstream analysis. We removed primers and low quality, archaeal and chloroplast reads in QIIME (Caporaso et al., 2010) and we eliminated chimeric sequences using UCHIME (Edgar et al., 2011). To cluster all the sequences in operational taxonomic units (OTUs) without using any arbitrary similarity cutoff, we used SWARM 2.0 algorithm (Mahé et al., 2014), applying a local clustering threshold of d=1, and the fastidious option to reduce the number of singletons and doubletons by grafting them onto more abundant clusters (Mahé et al., 2015). We further assigned taxonomy to the representative SWARMs by using SILVA 111 reference database (Quast et al., 2012) and the RDP classifier (Wang et al., 2007). We finally discarded all the OTUs that were represented by < 10 sequences and/or present in <10 samples and we randomly resampled the OTU table to generate an equal number of sequences per sample, based on the sample with the lowest number of reads (31,078 reads/sample in the Croche dataset and 30,900 reads/ sample in the 21 lakes dataset).

3.3.2 Clustering bacterial OTUs into categories of SpADs

In this study we have assessed the temporal dynamics of bacterial taxa that had been previously categorized based on their SpADs across boreal lakes (Niño-García *et al.* 2016). Details of the approach used to group bacterial taxa into categories of SpADs are presented in Niño-García *et al.* (2016). Briefly, we fit the abundance distribution of individual OTUs across the 198 lakes to statistical models using the Hartigans' dip test statistic (Maechler 2015, R Core Team 2014) and maximum likelihood estimation (fitdistrplus, Delignette-Muller and Dutang, 2014) and for each OTU we selected the statistical model with the lowest AIC value. This resulted in four different types of abundance distribution among all lake OTUs (normal-like, bimodal, logistic and lognormal), and thus we grouped all OTUs with a similar SpADs together into four different broad categories. To characterize the ecological features of the different categories of SpADs we compared their ranges in mean spatial abundance, spatial

occurrence and environmental breadth of their constituent OTUs. In addition, to further distinguish between categories influenced by hydrology-mediated dispersal or by environmental selection in lakes, we explored the dynamics of these OTUs across 188 additional streams and rivers (Strahler order 0 to 8) within the same boreal regions. This exercise showed that whereas normal-like and bimodal categories seemed to comprise the functional core of lake taxa, harboring bacteria that are responsive to local conditions, logistic and lognormal categories included mostly rare taxa whose presence seemed to be associated to dispersal from the watershed (Niño-García et al. 2016).

Since we had built the OTU table considering all those samples and the ones used in the present study together (198 lakes and 188 rivers in summer + annual series in Croche + spring and fall in the 21 lakes), the OTU names were common and here we could identify which of the OTUs from lake Croche or the 21 lakes had been spatially categorized as normal-like, bimodal, logistic or lognormal, and we grouped them as such.

3.3.3 Temporal dynamics of SpAD categories of bacteria along the annual cycle in lake Croche

We focus the analysis on OTUs detected in Lake Croche throughout the annual cycle that were also present in our previous large-scale study across lakes, and which had therefore already been classed into one of the four major categories of SpADs (Niño-García *et al.* 2016). We explored the temporal dynamics in abundance of each of these OTUs along the annual cycle in lake Croche. For each OTU within the four categories, we calculated the mean OTU abundance, occurrence, and environmental breadth over the annual cycle in lake Croche (based on temperature that was the variable with most pronounced changes), and we used these to derive a "temporal" average. In addition, we calculated the temporal turnover of OTUs classed within each of the categories, and compared all four properties with those we had previously calculated for the four categories of SpADs across lakes.

3.3.4 Exploring the presence of temporal shifters in lake Croche over the annual cycle

We explicitly assessed whether rare taxa belonging to the lognormal and logistic spatial categories, which had been classified as spatially "random" based on their SpADs, showed temporal recruitment along the annual cycle. We identified the OTUs that were categorized as logistic or lognormal in the spatial dataset, but that showed disproportionate changes in temporal abundance along the annual cycle in lake Croche. As these classes were characterized as rare in our spatial dataset, we consider an OTU to shift in abundance when it crossed a 0.1 % threshold of abundance (equivalent to 31 sequences at our rarefaction level) at least once during the whole annual cycle. From those, we further extracted OTUs whose mean spatial and temporal abundance were decoupled, focusing particularly on taxa that had systematically low spatial abundance, but which attained significantly higher abundance over time in lake Croche. We refer to these OTUs as "temporal shifters".

3.3.5 Extending the temporal analysis to boreal lakes

To expand this analysis beyond lake Croche, we explored the temporal abundance patterns of the taxa identified as "temporal shifters" across the 21 additional boreal lakes, which were part of the large-scale spatial summer study but for which we had data for two additional seasons (spring and fall). As in the analysis for lake Croche detailed above, OTUs were identified as temporal shifters if they had originally been classed in the lognormal and logistic spatial categories, and if they were rare during the summer in the 21 lakes but surpassed the 0.1 % threshold during spring and/or fall.

3.3.6 Statistical analyses

In order to assess the coherence between spatial and temporal ecological features of the different categories, we calculated the linear correlations between the corresponding mean annual abundance, occurrence environmental breadth and turnover. Spatial and temporal turnover were calculated based on Jaccard pair-wise dissimilarity (R betapart package, Baselga & Orme, 2012).

3.4 Results

The classification of OTUs in different categories of SpADs that we had carried out in a previous study (Niño-García *et al.* 2016), was only possible due to the large number of very different lakes included in the study (n=198). We could not repeat the same type of analysis over time in lake Croche because the relatively small number of time points (n=14) did not allow the same degree of temporal resolution. Therefore, rather than determining new categories of temporal dynamics and comparing these to the spatial categories, our approach was to identify those OTUs in lake Croche that had been previously detected in the large scale spatial analysis, and to assess whether their temporal dynamics matched the SpADs that we had previously observed for each of them. As mentioned before, this exercise was possible because the clustering of OTUs had been done using both the spatial and the temporal datasets, and thus the names were common.

We obtained a total of 466,170 sequences after quality control and rarefaction for the annual cycle in lake Croche (referred to as the "temporal" dataset hereafter), which clustered into 30,708 OTUs. The large majority of the OTUs (86%) that were present in lake Croche had already been detected in the previous large-scale survey across 198 boreal lakes (Niño-García et al. 2016), and had thus been classed into one of four categories of SpADs identified in that study. Only 4,162 OTUs (14% of Croche OTUs, 'Non-categorized' OTUs) were not found in the spatial dataset but these accounted for

a very small proportion of the sequences within the temporal dataset (<2%, Fig 3.1a). The majority of OTUs that we detected in lake Croche (66%) had been previously classed as either logistic or lognormal based on their SpADs, but together accounted for a relatively small proportion (29%) of all sequences detected in the lake over the annual cycle. On the other hand, most of the sequences (69%) detected in lake Croche

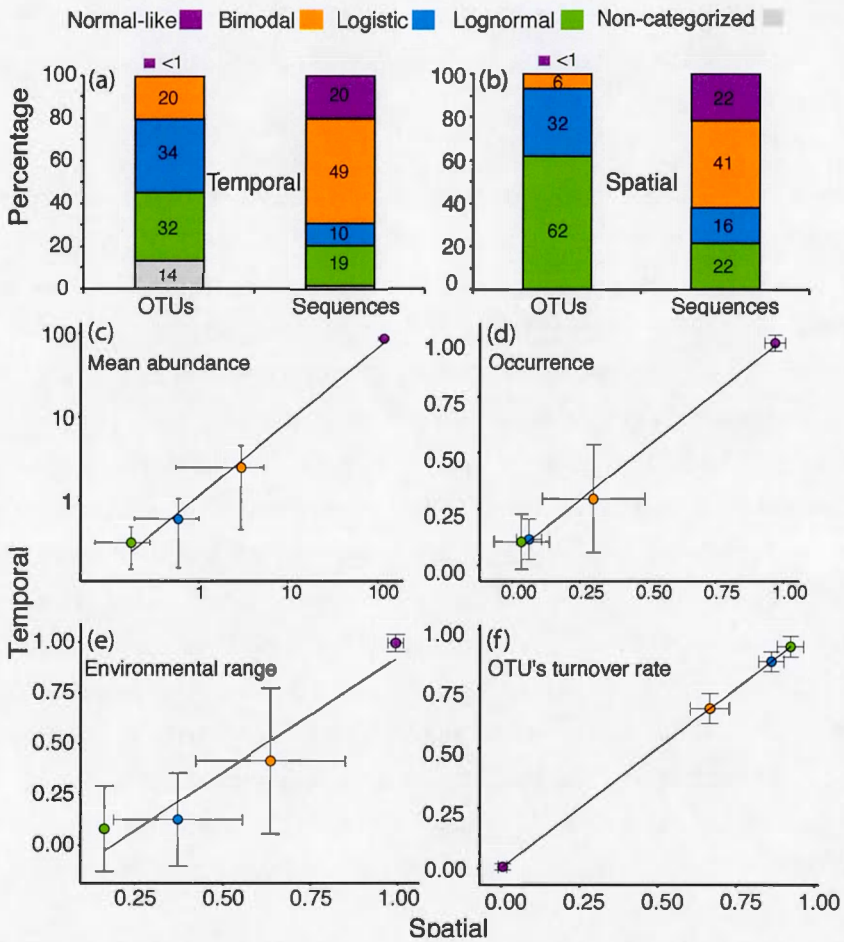


Figure 3.1. Exploration of the coherence between the temporal and spatial ecological properties for normal-like, bimodal, logistic and lognormal categories (represented by different colors): Contribution of each category to the total number of OTUs and associated sequences in lake Croche over time (a), and across the 198 boreal lakes (b), relationship between the average temporal and spatial mean abundances (c), occurrences (d), environmental ranges (estimated based on temperature in the case of Croche and pH in the spatial study) (e), and OTU turnover (f) for the different categories of SpAD. The dots represent the mean and the error bars the standard deviation associated to each mean.

were associated to a small proportion of OTUs (34%) that had been previously classed in the normal-like and bimodal categories based on their SpADs (Fig 3.1a, b). This overall distribution of OTUs and sequences detected over lake Croche annual cycle into the four categories of SpADs was remarkably similar to that observed over space for the ensemble of lakes, where we had found that over 94% of all OTUs found across all lakes classed into logistic and lognormal categories but accounted for relatively few sequences (38%), whereas a small number of OTUs classed into the normal and bimodal categories accounted for a large proportion (63%) of all sequences detected across lakes (Fig 3.1a,b).

There was a very strong correlation between the averaged spatial and temporal mean OTU abundances, occurrences, environmental ranges and the OTU turnover, suggesting a strong overall coherence between the ecological properties of our four categories on temporal and spatial scales (Fig 3.1c-f). For example, spatially normal-like OTUs seemed to be also very abundant and ubiquitous over the year in lake Croche, and most logistic and lognormal OTUs were extremely rare and present only in a few occasions annually (Fig. 3.1c-f). Although the number of OTUs associated to the bimodal and logistic categories in lake Croche was slightly below the average values in the temporal compared to the spatial dataset (Fig. 3.2a), in general the temporal variability in the number of OTUs classed within each of the four spatial categories, and the number of associated sequences, fell well within the spatial range of variation in these parameters observed across the 198 lakes (Fig. 3.2a,b). These results confirm that the OTUs that appeared to be selected by lake environmental conditions spatially (normal-like and bimodal SpAD categories), and therefore making up the core component of bacterial communities across lakes, corresponded to those that dominated the lake Croche bacterial community over time and were reactive to environmental changes along an annual cycle.

We were particularly interested in assessing whether any of the rare OTUs classed in logistic and lognormal categories, whose distribution was considered as random and hydrology-mediated (Niño-García *et al.* 2016), showed active recruitment along the annual cycle in lake Croche. We found that the vast majority (ca. 99%) of these OTUs were also systematically rare in Lake Croche, never exceeding an arbitrary abundance

threshold of 0.1% at any point of the annual cycle, and never showing any kind of clear seasonal dynamics, confirming that most spatially rare taxa remain temporally rare as well. We detected a small subset of spatially rare taxa, however, which did show seasonal transitions from rare to abundant (details not shown), composed of 115 logistic OTUs and 206 lognormal OTUs that exceeded the 0.1% abundance threshold at least once over the course of the annual cycle. It is interesting to note that for over half of these OTUs there was a strong relationship between their average spatial and temporal abundances, which suggests that there was coherence between their spatial and temporal patterns of distribution (grey dots in Figs. 3.3a,c, which were indeed outliers both in the spatial and the temporal dataset). Only 43 and 36 OTUs,

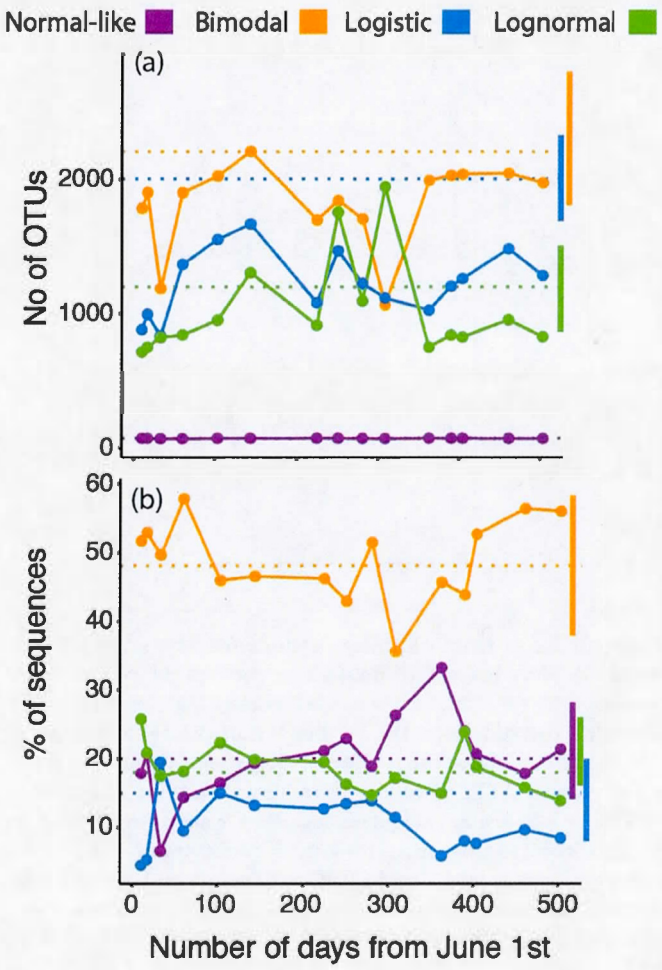


Figure 3.2. Temporal patterns in the number of OTUs (a) and the percentage of sequences (b) associated to normal-like, bimodal, logistic and lognormal categories along the annual cycle in lake Croche, in relation to their respective means (dotted horizontal lines) and ranges of variation (vertical bars) derived from the large-scale spatial dataset (198 lakes).

respectively, showed a clear decoupling between their respective spatial and temporal mean relative abundances (vertical lines in Figs 3.3a,c), having much higher mean temporal abundances within lake Croche than their average summer abundance across all lakes. These temporal shifters thus represent OTUs that were spatially rare during the summer across all lakes, but that recruited to higher abundances at other times of the year in lake Croche. Although these shifter OTUs showed peaks at different times of the year (see examples of individual OTUs in Figs. 3.3b,d), their aggregated dynamics (black line in Figs. 3.3b,d) showed a prevalence of such shifts in abundance in times of the year different from summer, many during winter (Fig 3.3b, d). It is

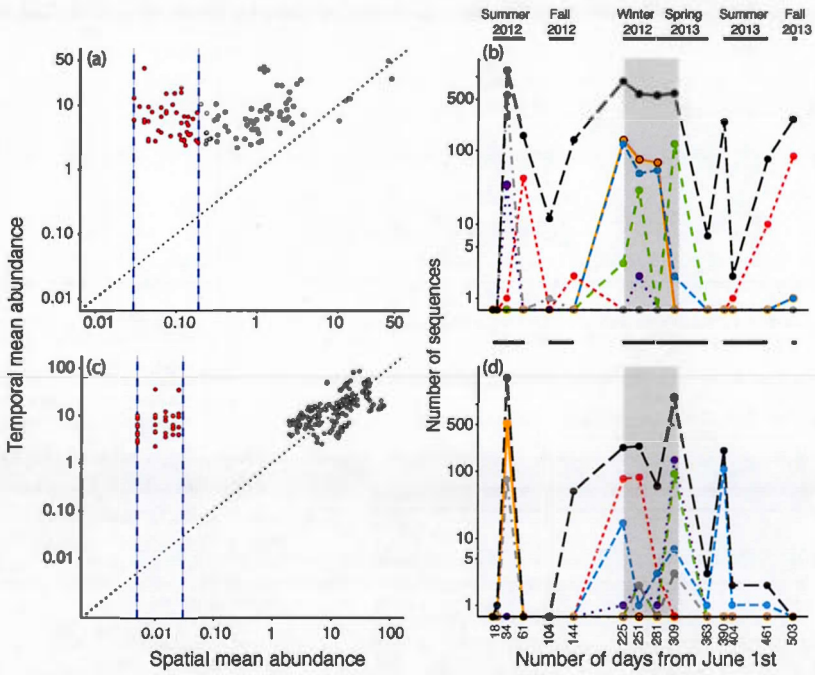


Figure 3.3. Relationship between the spatial and temporal mean abundance of logistic (a) and lognormal (c) OTUs that crossed the 0.1% abundance threshold along the annual cycle in lake Croche. Temporal shifters (red circles) are OTUs that had a disproportionately high relative abundance in lake Croche over the annual cycle (position over the 1:1 black dotted line) and a mean spatial abundance that was within the range used to define it either as logistic or lognormal (vertical blue dashed lines). Grey circles correspond OTUs discarded as temporal shifters, which were not within the ranges of typical logistic and lognormal categories (i.e., they were outliers both in the temporal and the spatial surveys). Examples of temporal dynamics of representative individual logistic (b) and lognormal (d) temporal shifters in lake Croche. Different colors and/or line types correspond to individual OTUs, and the black dashed line represents their pooled sequences. Horizontal black lines on top of the plots show the corresponding sampling seasons and the vertical shade represent the ice-covered period.

clear that this group of temporally shifting OTUs is not well categorized only based on their spatial dynamics across lakes in summer, yet it is important to note that they comprise a negligible fraction of the total number of OTUs identified within the logistic and lognormal rare categories (0.5% and 0.7%, respectively). Together, these results demonstrate that only a very small proportion of the spatially rare, random OTUs (i.e. logistic and lognormal) appear to recruit temporally and ever become part of the core component of the community in lake Croche.

In order to assess the generality of these findings, we explored the distribution and prevalence of temporal lake Croche shifters in an additional dataset of 21 lakes for which we had samples for three distinct seasons (summer –also included in the spatial dataset-, spring and fall). Of the 79 lognormal or logistic temporal shifters previously identified in Lake Croche (Figs. 3.3a,c), 72 were also present in at least one of the 21 lakes (mean per lake = 4, range 0 – 21) but they were all consistently below the 0.1 % abundance threshold in all three seasons, and therefore did not behave as temporal shifters in these lakes. This suggests that the logistic and lognormal taxa that shift temporally may be highly lake specific, as opposed to the normal and bimodal taxa, which are widespread across all lakes, and also in time within a given lake.

In order to further test this hypothesis, we independently identified temporal shifters in the 21-lake dataset by selecting the OTUs whose abundance was below the 0.1% abundance threshold in summer (because this is when their spatial categorization was made in Niño-García et al. 2016), but which became abundant in spring and/or fall. This procedure recovered a total of 127 shifter taxa belonging to the logistic and lognormal categories (Table 3.1). Interestingly, the mean number of shifter OTUs per lake was similar to that found in lake Croche (Table 3.1), suggesting that in lakes in general there may exist a small pool of taxa that appear to be spatially rare, but which spike at very specific times of the year and which are therefore reactive to certain lake environmental conditions.

3.5 Discussion

Modern genomic approaches have revealed that aquatic and terrestrial bacterial

		Number of OTUs	% of Sequences
Lake Croche	**Total**	79	-
	Logistic	43	0.01
	Lognormal	36	0.01
Logistic in 21 lakes dataset	**Total**	60	-
	Mean	13	0.03
	Range	0 - 33	0 -2
Logistic in 21 lakes dataset	**Total**	67	-
	Mean	25	0.2
	Range	8 - 45	0.1 -3.5

Table 3.1 Number of OTUs and percentage of sequences categorized as temporal shifters in lake Croche and in the dataset of 21 lakes for which we had seasonal data (see text).

communities are composed of thousands of taxa, the vast majority of which are extremely rare, both in abundance and in occurrence across systems (Curtis et al., 2002; Lynch and Neufeld, 2015; Sloan et al., 2006). The structure of these communities results from a combination of active growth within the ecosystem, and dispersal from outside the system (Pedrós-Alió, 2006; 2012) but one of the major contemporary challenges in microbial ecology has been to discriminate the core from the random components of bacterial communities (Caporaso et al., 2011; Saunders et al., 2015; Staley et al., 2014). In a companion paper we addressed this fundamental issue explicitly for lake bacterial communities (Niño-García et al. 2016), known to be strongly structured by both environmental selection and immigration from adjacent ecosystems (Crump et al. 2012; Ruiz-González et al 2015; Niño-García et al. 2016). In that previous study we had determined the spatial abundance distribution of individual bacterial OTUs (determined on the basis of deep sequencing of 16S rDNA) across 198 widely different boreal lakes in Québec, and we used these to group taxa into categories of SpADs. This partition of the structure of lake bacterial communities into four groups of taxa with different SpADs suggested that lake communities are composed of a core of taxa whose distribution is linked to active growth and in-lake environmental selection (normal-like and bimodal categories of SpAD), and an enormous fraction of rare bacteria (94% total OTUs) whose presence seems random and linked to hydrology-mediated transport (logistic and lognormal categories, Niño-García et al. 2016).

A major outstanding question that emerged from the above studies is whether this

functional discrimination of lake bacterial taxa based on their mid-summer spatial distribution across boreal lakes may have a temporal dimension, wherein for example taxa that are spatially rare and apparently "random" may in fact be reactive at other times of the year, or taxa that appear to be widespread and consistently abundant disappear in particular seasons. In this study we explicitly addressed this question, by assessing the temporal dynamics of those taxa that had been previously classified on the basis of their large-scale SpADs across lakes. The number of temporal samples that we collected from lake Croche did not allow us to model the abundance/occurrence distributions over time as we had done across lakes (14 temporal samples compared to 198 spatial samples), so we based our analysis on identifying OTUs in lake Croche that had previously been classified into one of the four categories of SpAD. This allowed us to study whether the ecological features attributed to these taxa based on their spatial dynamics do apply also when the temporal scale is considered, and therefore, how accurately this spatial categorization of taxa reflects intrinsic properties or capacities of individual bacterial OTUs. Our results demonstrate a high degree of coherence between the ecological features inferred from the patterns of spatial and temporal distribution of bacterial taxa, in particular concerning the distinction between taxa that are reactive to environmental conditions versus those that are non-reactive and whose presence in lakes appears to be accidental (Fig 3.1).

In our previous study we had established that all lake communities contain a small group of normal-like OTUs that are both consistently abundant and widespread across lakes, which appear to be highly tolerant to wide environmental gradients (Niño–García et al. 2016). Here we show that most of those OTUs that were previously categorized as normal-like were also present in lake Croche year round, and accounted for a large fraction of the total sequences detected throughout the annual cycle (Figs. 3.1a,b,c,d, 3.2b). The temporal dynamics of this group of phylotypes supports the previous conclusion based on their SpADs that this category represents ubiquitous taxa possibly with high intrinsic growth rates that are tolerant to wide environmental ranges (Lennon et al., 2012; Mou et al., 2008; Newton et al., 2011; Székely and Langenheder, 2014), and therefore present not only across all lakes but also throughout the year within a given lake (Fig 3.2b). The second major category of bacterial phylotypes that we had identified from their SpADs were the bimodals, composed of OTUs with much more

variable abundances and occurrences across lakes than the OTUs within the normal-like category, and in general with a narrower range of environmental preference, yet collectively accounting for the largest proportion of the sequences in the spatial dataset (Niño-García et al. 2016, Fig. 3.1b). These features were essentially replicated in time over the annual cycle in lake Croche (Fig. 3.1a), and OTUs belonging to the spatial bimodal category had strong seasonal fluctuations, as reflected by their high temporal turnover, and which collectively accounted for a large fraction of the total number of sequences detected in the lake (Fig. 3.1c,e, Fig 3.2b), as was the case across the 198 lakes. Thus, whereas normal-like OTUs are consistently present and abundant in lake bacterioplankton communities over space and time, bimodal OTUs represent bacteria that, while also reactive to lake conditions, tend to cycle between the rare and abundant fractions of the community possibly due to their narrower range of preferences and/ or tolerances (Lennon et al., 2012; Lynch and Neufeld, 2015; Salcher, 2013; Shade and Gilbert, 2015). In any case, the temporal and spatial patterns converge to suggest that these two categories likely harbor the bulk of the active component of lake bacterioplankton communities, thus likely representing the functional core of these communities as proposed by Shade and Handelsman (2012) and Saunders et al. (2016).

In our previous study, the vast majority of the rare taxa fell into two distinct categories of SpADs (logistic and lognormal) based on their cross-lake changes in abundance, and their patterns along the associated fluvial networks suggested that their presence in lakes was mostly driven by dispersal from the watershed (Niño-García et al. 2016). We had thus hypothesized that most of these taxa are random within lakes, likely due to a combination of downstream transport, persistence and dilution. In this regard, we expected that the temporal patterns of these "random" taxa should be influenced by temporal variations in hydrological loading within the lake, and thus for example we would expect a higher proportion of logistic and lognormal OTUs in times of high water inputs, such as during snow melt or after strong storms. We did not detect any clear seasonal pattern associated to such events in these two spatial categories (details not shown), but their aggregated ecological features were very similar in both the spatial and temporal datasets: they represented the majority of OTUs found along the annual cycle in Lake Croche (Fig 3.1), yet accounted for fewer than half of the total sequences (Fig 3.2b), had very low average abundance and an extremely patchy

temporal occurrence, a pattern that is very similar to what we had previously found spatially across lakes (Niño-García et al. 2016). There was therefore an overall tight coherence between the spatial and temporal dynamics of these rare taxa as a whole, and no clear evidence that the inclusion of a temporal dimension alters the conclusion that their presence in lakes is mostly random.

It is possible, however, that individual taxa within this overall pool of rare, apparently non-reactive bacteria may recruit at specific times of the year (Alonso-Sáez, 2014; Shade and Gilbert, 2015; Shade et al., 2014), and this might not be reflected in the average temporal dynamics of the group as a whole. Interestingly, when exploring the individual temporal patterns of all the OTUs classed in the lognormal and logistic categories, we found a very small number of OTUs (79) that were spatially rare but which became temporarily abundant in lake Croche (temporal shifters, Fig 3.3a,c). The overwhelming majority of lognormal and logistic OTUs remained below the abundance threshold we set (0.1%) and were patchily distributed throughout the annual cycle, again confirming that these taxa are mostly unreactive to lake environmental conditions (Niño-García et al. 2016). It is interesting to note that the temporal patterns were remarkably similar among some of the temporal shifters that we identified (Fig. 3.3c,d), suggesting either strong eco-physiological similarities between these taxa (Barberán et al., 2011; Eiler, Heinrich, & Bertilsson, 2011; Steele et al., 2011; Williams, Howe, & Hofmockel, 2014), or a common source. Most of these shifters tended to peak in winter, suggesting that both temporal changes in local environmental conditions, and unusual loading events may underlie the temporal spikes of these taxa (Fig 3.3b,d). This dynamics would be analogous to that of the conditionally rare taxa identified by Shade et al. (2012), which have been shown to contribute disproportionately to temporal changes in microbial communities, but in lake Croche these potential shifters comprise a very small fraction of the vast pool of rare bacteria. Rather, the vast majority of the taxa that showed strong temporal variations in abundance and occurrence were spatially bimodal taxa, confirming that the reactivity of the taxa in this category to lake environmental conditions is expressed both spatially across lakes, and temporally within a given lake (Fig 3.1 c,d).

The relatively small number of rare (logistic and lognormal) taxa that we identified

as temporal shifters is rather surprising, and may be the result of having considered only one lake for this temporal analysis. If such pool of rare OTUs that can potentially recruit at specific times over an annual cycle were very site-specific, the size of this reactive pool of spatially rare taxa should increase progressively with an increasing number of lakes considered. However, when we repeated the analysis of the temporal patterns of individual logistic and lognormal OTUs for the additional 21 lakes for which we had seasonal data, we recovered only 1.6 fold more temporally shifters than when considering lake Croche alone (127 vs. 79, Table 3.1). Although we acknowledge the limited temporal resolution of this multi-lake dataset, especially considering that many shifters in lake Croche recruited during winter (Fig 3.3d), it does not seem that increasing the number of lakes proportionately increased the apparent size of this pool of rare, temporally reactive bacteria. In addition, we found that whereas most of the temporal shifters that we identified in lake Croche were detected in the 21-lake dataset, these were always below the 0.1% threshold of abundance and at relative low occurrences (ca. 20 %) in the 21 lakes. The spatial turnover of these temporal shifters OTU across the 21 lakes (spatial OTUs turnover 0.2 -0.7) suggests an incomplete replacement of temporal shifters between lakes (i.e., many are shared between lakes), but their ability to recruit seems system-dependent because most of these shifters peaked in only one or two lakes. It is interesting to note that at a broad phylogenetic level these spatially rare temporal shifters basically resemble the composition of the ensemble of rare logistic and lognormal categories (Fig. 3.4a), yet at a finer phylogenetic resolution, there are some obvious differences, with *Sphingobacteria* and *Flavobacteria* Classes being disproportionately represented (Fig. 3.4b). This may indicate eco-physiological commonalities between these taxa that should be further explored.

We acknowledge that the 0.1% abundance threshold that we used to identify these logistic or lognormal temporal shifters is arbitrary, although it has frequently been used in previous papers to separate rare vs. abundant taxa (Fuhrman, 2009; Pedrós-Alió, 2006; Vergin et al., 2013). Lowering or increasing this threshold, however, will not change the overall observation that the vast majority of rare OTUs classed within the lognormal and logistic spatial categories do not show any discernible temporal pattern in lakes. Our results thus support our previous observation that the lake "rare biosphere" is mostly composed of OTUs whose presence is accidental and linked to

hydrologic dispersal (Niño-García et al. 2016).

Overall, we show that the ecological properties derived from the patterns of spatial abundance distribution of bacteria across space seem to be ecologically robust and inherent to taxa since their dynamics over time is largely what we expected from each of the 4 categories of OTUs. In general, taxa that are able to grow at a very

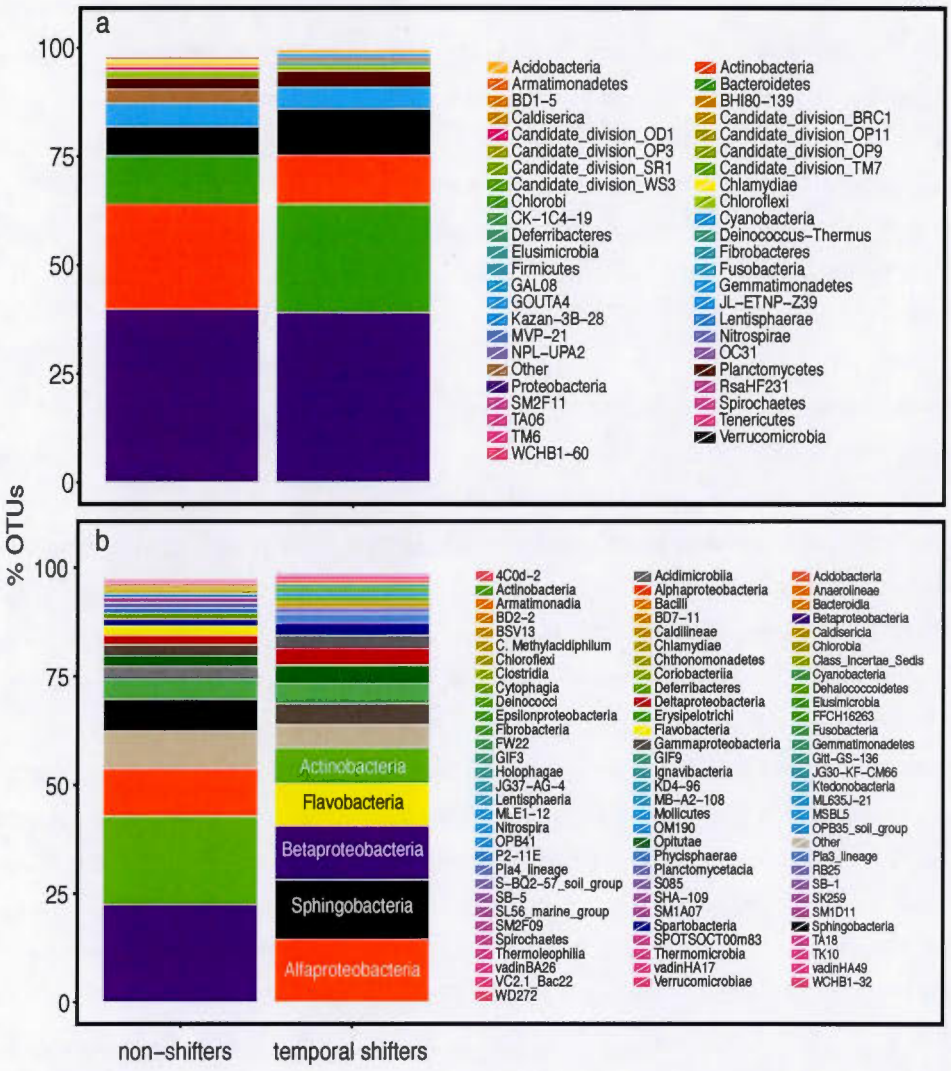


Figure 3.4. The taxonomic composition of non-shifters and temporal shifters at the Phylum (a) and Class taxonomic levels (b).

broad range of lake environmental conditions (normal-like) are also able to tolerate temporal changes in these conditions and dominate year long. Taxa that are restricted to a more limited set of environmentally similar lakes will also tend to be prevalent at more specific times of the year (i.e. bimodal), and bacteria whose spatial patterns are the result of downstream dispersal and persistence (lognormal and logistic) also tend to be unreactive at the temporal scale. Interestingly, the remarkable coherence in the contribution of each of these four categories to the structure of the communities suggests that this must be a recurrent property of boreal freshwater bacterioplankton assemblages. This implies that the reactive core of lake bacterioplankton communities is mainly composed by an adaptive group of normal-like and bimodal OTUs that responds to spatial and temporal environmental gradients. In contrast, downstream-dispersal processes drive the distribution of the vast majority of accidental bacteria, which do ppear to be reactive to ambient conditions and likely play a limited role in community processes.

Our results further support the view that biological communities can be partitioned into subsets of taxa whose different abundance distributions result either from stochastic processes related with dispersal or from niche-driven mechanisms associated to selection by local conditions that together influence the shape of the species abundance distributions (Magurran and Henderson, 2003; Ulrich and Zalewski, 2006; Ulrich and Ollik, 2004; Hanski, 1982). The large role of dispersal on structuring freshwater bacterioplankton communities and its variation across the landscape (Niño-García *et al.*, 2016; Ruiz-González *et al.* 2015) suggest that the balance between the core and the random components of bacterioplankton communities likely changes along the aquatic continuum: in fast-flowing headwater streams, which are strongly subjected to inoculation of microbes from soils (Ruiz-González et al 2015), the contribution of such accidental taxa may be relatively more important than downstream in the networks, whereas the core component will make up a larger fraction of the communities in systems with higher water residence times. The coherence between the spatial and temporal properties measured here for the different categories of SpADs, together with the remarkably small number of temporal shifters, indicate that similar processes underpin spatial and temporal abundance distributions of bacterioplankton taxa in lakes, and that bacterial rarity on these ecosystems is largely the result of hydrologic

mediated transport of unreactive bacterial taxa.

3.6 Acknowledgements

We thank Annick St-Pierre, Alice Parkes and the whole CarBBAS team for their contribution to the field and laboratory components of this research. We also thank Carolina García Cháves for collecting the samples of lake Croche, and to Nicolas Fortin St-Gelais for critical comments. This study is part of the program of the Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBAS) Industrial Research Chair, co-funded by the Natural Science and Engineering Research Council of Canada (NSERC) and Hydro-Quebec.

	21-lakes dataset		198-lakes dataset	
Variable	**Minimum**	**Maximum**	**Minimum**	**Maximum**
pH	5.56	9.22	4.04	8.95
DOC (mg/L)	3.96	20.25	1.064	39.67
cDOM (m^{-1})	0.01	8.64	0.01	20.04
TN (mg/L)	0.4	0.89	0.1	0.91
TP (µg/L)	2.88	117.6	2.58	153
TEMP (C)	7.5	24.06	10.75	24.7
CHL (µg/L)	0.39	10.24	0.14	36.05
WRT(days)	17	10523	1.38	10523
Zmax (m)	0.7	70	0.5	78
Secchi (m)	0.1	6.5	0.1	9.5

Table 3.2 Range of main limnological and landscape variables measured in the 21-lakes and 198-lakes dataset. Whereas for 198 lakes the ranges correspond to a summer snapshot, the 21-lakes dataset includes the variation in spring, summer and fall.

GENERAL CONCLUSIONS

4.1 Main contributions of this thesis

The central topic of this thesis focuses on the development of our understanding of how biogeographic factors and processes influence the composition of freshwater bacterioplankton communities in lakes. We have used a large-scale approach that involved extensive sampling of freshwater ecosystems that cover wide ranges in geographic, hydrologic and environmental conditions, a scale rarely found in the literature. This approach allowed us to assesses the interplay between different large-scale biogeographic processes on the distribution of environmentally active versus non-reactive, random taxa, which is a greatly understudied aspect of bacterioplankton community structure. This thesis thus contributes not only to our understanding of the factors that shape aquatic bacterial communities, but also those that describe the origin, structure, and ecological role of the vast rare biosphere that populates all aquatic bacterial assemblages. Specifically, the following key points emerge from the chapters of this thesis:

1) The thesis demonstrates that the interplay between landscape and local ecological properties shapes the composition of bacterioplankton communities in boreal freshwater ecosystems. The combined influence of hydrology, network position, and water chemistry shape the directional structuring of local aquatic communities from highly diverse assemblages in headwater streams, to larger rivers and lakes dominated by fewer taxa, regardless of geographic location. Although environmental sorting by local conditions and mass effects both occur across the whole aquatic network, hydrology and network position modulate their relative influence on taxonomic composition at any given location by defining both the time frame for bacterial growth and the composition of the immigrant pool. Collectively, in combination with the lack of evidence for dispersal limitation, these findings highlight both the significance of interconnectivity of aquatic ecosystems across the landscape, and the role of the origin of the immigrant pool of bacteria entering aquatic networks in shaping the composition

of local bacterial communities.

2) Boreal lake bacterioplankton communities are made up of a small core of reactive taxa and a very large fraction of bacteria whose presence in lakes is the result of passive transport, but distinguishing these two components is a major challenge. We used the large-scale patterns of abundance and occurrence across lakes to discriminate between these two components of lake bacterioplankton. Interestingly, all bacteria taxa fit within a small set of discrete, categories of spatial abundance distributions (SpADs) that show dramatic differences in their ranges of abundance, occurrence, and environmental tolerance regardless of taxonomic identity, as well as differences in local dynamics within riverine networks. (Chapter II). Two of these categories of SpADs (normal-like and bimodal) grouped the most abundant and ubiquitous taxa, which were clearly linked to active in-lake selection and recruitment. The other two categories of spatial behaviour (logistic and lognormal) grouped mostly rare bacteria whose patterns within the aquatic network clearly suggested passive transport, driven by the flow of water along the aquatic continuum. The presence of the taxa belonging to these two latter categories, which collectively group most of the OTUs found across lakes, appears to be mostly random.

3) The presence of a small group of ubiquitous rare but actively selected taxa in lakes (Chapter II), and the very small number of rare taxa that show large temporal shifts in abundance within lakes (temporal shifters, Chapter III), suggest the existence of bacteria with adaptive strategies of survival and persistence within the "rare biosphere". The vast majority of rare taxa that were catalogued as random based on their large-scale summer SpADs did not show any kind of reactivity to lake conditions over time. There was thus a remarkable coherence between the spatial distributions and the temporal dynamics of the categories mentioned above (Chapter III), suggesting that the presence in lakes of the vast majority of rare bacterial taxa is random and driven by passive transport processes, whereas the core, active component of these lake communities is dominated by an adaptive group of normal-like and bimodal OTUs that respond to spatial and temporal environmental gradients through strong shifts in their relative abundances.

4.2 Novelties and main implications

The sampling design implemented throughout this thesis allowed for the first empirical assessment of the interplay between geography, landscape features and local environmental conditions on bacterial taxonomic composition across large-scale boreal freshwater networks. The main results establish a strong interaction between hydrology and water chemistry in structuring local freshwater bacterioplankton communities but they also show that the relative influence of mass effects and environmental sorting change along the aquatic continuum. Thus, these findings stress that interpretations based on partial fragments of the complex biogeographic reality described above may lead to a biased understanding of the factors shaping freshwater bacterioplankton communities over large spatial scales and, therefore, of the relative influence of biogeographic processes on the assembly of local bacterial communities. The analyses presented in this thesis (Chapter I) offer a general framework to contextualize previous and future results and this, beyond the boreal region. This framework allows to effectively integrate fragmented and often ambiguous or contradictory conclusions that exists in the current literature concerning the main drivers of the spatial variability in freshwater bacterial community composition.

The novelty of these results lies in the demonstration that, regardless of their environmental preferences, lake bacterioplankton communities are composed of taxa that can be grouped into a few discrete categories of SpADs that effectively discriminate core-active versus passive-transport random taxa across lakes. This approach is based on a large-scale comparative perspective that accounts not only for the distribution of taxa across lakes, but also of their actual movement and transport within the associated aquatic networks. Contrary to other studies that have assessed the activity of individual taxa within a specific ecosystem, our approach integrates the influence of large-scale biogeographic process shaping bacterial community composition. The resulting spatial patterns are used to infer aggregated ecological features that result from the combination of intrinsic eco-physiological traits of taxa, which in turn determine their potential growth rates, environmental breadth, and capacity to persist under unfavorable conditions.

Finally, this thesis includes the first empirical estimate of how these distinct functional categories are distributed across lakes and over time, and presents evidence for the numerical relevance of different types of ecological strategies within the "rare biosphere". Since the vast majority of rare taxa in lakes seem to be random and driven by passive transport processes, this "rare biospheres" seems to play a limited role in community processes, and more importantly, provides little insight into lake functional niches and responses to environmental or biotic interactions. As this vast rare component is highly influenced by hydrologic transport, it may also represent the network history that precedes local community assembly in lakes. As such, the rare component of the bacterioplankton community might not reflect current lake conditions as much as it might describe the origin of its resident bacterial communities. Accordingly, this fraction of bacterial communities might provide information on the position and degree of connectivity of a given lake within the hydrologic continuum.

REFERENCES

Aanderud, Z. T., Jones, S. E., Fierer, N., & Lennon, J. T. (2015). Resuscitation of the rare biosphere contributes to pulses of ecosystem activity. Frontiers in Microbiology, 6, 217.

Adler, P.B. and Lauenroth, W.K. (2003). The power of time: spatiotemporal scaling of species diversity. Ecology Letters 6, 749– 756.

Adler, P.B., White, E.P., Lauenroth, W.K., Kaufman, D.M., Rassweiler, A. & Rusak, J.A. (2005). Evidence for a general species-time-area relationship. Ecology, 86, 2032–2039.

Adams, H. E., Crump, B. C., & Kling, G. W. (2014). Metacommunity dynamics of bacteria in an arctic lake: the impact of species sorting and mass effects on bacterial production and biogeography. Frontiers in Microbiology, 5, 82.

Alonso-Sáez, L. (2014). Winter bloom of a rare betaproteobacterium in the Arctic Ocean. Frontiers in Microbiology, 1–9.

Alonso-Sáez, L., & Gasol, J. M. (2007). Seasonal Variations in the Contributions of Different Bacterial Groups to the Uptake of Low-Molecular-Weight Compounds in Northwestern Mediterranean Coastal Waters. Applied and Environmental Microbiology, 73(11), 3528–3535.

Alonso-Sáez, L., Díaz-Pérez, L., & Morán, X. A. G. (2015). The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. Environmental Microbiology, 17(10), 3766–3780.

Alonso-Sáez, L., Zeder, M., Harding, T., Pernthaler, J., Lovejoy, C., Bertilsson, S., & Pedrós-Alió, C. (2014). Winter bloom of a rare betaproteobacterium in the Arctic Ocean. Frontiers in Microbiology, 5, 425.

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. Biometrics, 62(1), 245–253.

Anderson, M. J. (2011). A new method for non-parametric multivariate analysis of variance. Marti J. Anderson, 1–23.

Baltar, F., Palovaara, J., Vila-Costa, M., Salazar, G., Calvo, E., Pelejero, C., et al. (2015). Response of rare, common and abundant bacterioplankton to anthropogenic perturbations in a Mediterranean coastal site. FEMS Microbiology Ecology, 91(6).

Barberán, A. (2014). The microbial contribution to macroecology, 1–8.

Barberán, A., & Casamayor, E. O. (2010). Global phylogenetic community structure and β-diversity patterns in surface bacterioplankton metacommunities. Aquatic Microbial Ecology, 59, 1–10.

Barberán, A., Bates, S. T., Casamayor, E. O., & Fierer, N. (2011). Using network analysis to explore co-occurrence patterns in soil microbial communities. The ISME Journal, 1–9.

Baselga, A., & Orme, C. D. L. (2012). betapart: an R package for the study of beta diversity. Methods in Ecology and Evolution, 3(5), 808–812.

Besemer K, Singer G, Quince C, Bertuzzo E, Sloan W, Battin TJ. (2013). Headwaters are critical reservoirs of microbial diversity for fluvial networks. *Procedings of the Royal Society B* 280: 20131760.

Borcard, D., Gillet, F., & Legendre, P. (2011). Numerical Ecology with R. Springer Science & Business Media.

Campbell, B. J., Yu, L., Heidelberg, J. F., & Kirchman, D. L. (2011). Activity of abundant and rare bacteria in a coastal ocean. Proceedings of the National Academy of Sciences of the United States of America, 108(31), 12776–12781.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7(5), 335–336.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal, 6(8), 1621–1624.

Caporaso, J. G., Paszkiewicz, K., Field, D., Knight, R., & Gilbert, J. A. (2011). The Western English Channel contains a persistent microbial seed bank. The ISME Journal, 6(6), 1089–1093.

Carignan, R., Planas, D., & Vis, C. (2000). Planktonic production and respiration in oligotrophic Shield lakes. Limnology and Oceanography.

Carini, P., Marsden, P. J., Leff, J. W., Morgan, E. E., Strickland, M. S., & Fierer, N. (2016). Relic DNA is abundant in soil and obscures estimates of soil microbial diversity.

Comte, J., Lovejoy, C., Crevecoeur, S., & Vincent, W. F. (2016). Co-occurrence patterns in aquatic bacterial communities across changing permafrost landscapes. Biogeosciences, 13(1), 175–190.

Cotner, J. B., & Biddanda, B. A. (2002). Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems. Ecosystems, 5(2), 105–121.

Cottrell, MT, del Giorgio, PA, Kirchman, DL. (2015). Biogeography of globally distributed bacteria in temperate and boreal Québec lakes as revealed by tag pyrosequencing of 16S rRNA genes. Aquatic Microbial Ecology. doi: 10.3354/ame01776

Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. Ecology Letters, 8(11), 1175–1182.

Cottrell, M. T., & Kirchman, D. L. (2000). Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low-and high-molecular-weight dissolved organic Matter. Applied and Environmental Microbiology, 66(4), 1692–1697

Crump, B. C., Adams, H. E., Hobbie, J. E., & Kling, G. W. (2007). Biogeography of Bacterioplankton in Lakes and Streams of an Arctic Tundra Catchment. Ecology, 88(6), 1365–1378.

Crump, B. C., Amaral-Zettler, L. A., & Kling, G. W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. The ISME Journal, 6(9), 1629–1639. http://doi.org/10.1038/ismej.2012.9

Crump, B. C., Hopkinson, C. S., Sogin, M. L., & Hobbie, J. E. (2004). Microbial Biogeography along an Estuarine Salinity Gradient: Combined Influences of Bacterial Growth and Residence Time. Applied and Environmental Microbiology, 70(3), 1494–1505.

Curtis, T. P., Sloan, W. T., & Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences of the United States of America, 99(16), 10494–10499.

del Giorgio, P. A., & Cole, J. J. (1998). Bacterial growth efficiency in natural aquatic systems. Annual Review of Ecology and Systematics, 29, 503-541.

del Giorgio, P. A., & le B Williams, P. J. (2005). Respiration in Aquatic Ecosystems. Oxford University Press, USA.

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software, 64(4), 1-34.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics (Oxford, England), 27(16), 2194–2200.

Eiler, A., Heinrich, F., & Bertilsson, S. (2011). Coherent dynamics and association networks among lake bacterioplankton taxa. The ISME Journal, 6(2), 330–342.

Evans, S. E., & Wallenstein, M. D. (2013). Climate change alters ecological strategies of soil bacteria. Ecology Letters, 17(2), 155–164.

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science, 320(5879), 1034–1039.

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences of the United States of America, 103(3), 626–631.

Fierer, N., Bradford, M. A., & Jackson, R. B. (2007a). Toward an ecological classification of soil bacteria. Ecology, 88(6), 1354–1364.

Fierer, N., Morse, J. L., Berthrong, S. T., Bernhardt, E. S., & Jackson, R. B. (2007b). Environmental controls on the landscape-scale biogeography of stream bacterial communities. Ecology, 88(9), 2162–2173.

Fujii M, Kojima H, Iwata T, Urabe J, Fukui M. (2012). Dissolved organic carbon as

major environmental factor affecting bacterioplankton communities in mountain lakes of Eastern Japan. *Microbial Ecology* 63:496–508.

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. Nature, 459(7244), 193–199.

Galand, P. E., Casamayor, E. O., Kirchman, D. L., & Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. Proceedings of the National Academy of Sciences, 106(52), 22427–22432.

Gasol, J. M., & del Giorgio, P. A. (1997). Biomass distribution in marine planktonic communities. Limnology and Oceanography, 42(6)1353-1363.

Geobase. (2009. Land Cover, Circa 200-Vector

Gibbons, S. M., Caporaso, J. G., Pirrung, M., Field, D., Knight, R., & Gilbert, J. A. (2013). Evidence for a persistent microbial seed bank throughout the global ocean. Proceedings of the National Academy of Sciences of the United States of America, 110(12), 4651–4655.

Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrück, L., Reeder, J., Temperton, B., et al. (2012). Defining seasonal marine microbial community dynamics. The ISME Journal, 6(2), 298–308.

Guo, Q., Brown, J.H. and Valone, R. (2000). Abundance and distribution of desert annuals: are spatial and temporal patterns related? Journal of Ecology, 88, 551–560.

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., & Martiny, J. B. H. (2012). Beyond biogeographic patterns:processes shaping the microbiallandscape. Nature Reviews Microbiology, 10(7), 497–506.

Hanski, I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. Oikos, 38, 210-221.

Hanski, I. and Gyllenberg, M. (1993). Two general metapopulation models and the core-satellite species hypothesis. American Naturalist, 142, 17–41.

Harte, J., Conlisk, E., Ostling, A., Green, J.L. & Smith, A.B. (2005). A theory of spatial

structure in ecological communities at multiple spatial scales. Ecological Monographs, 75, 179–197.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25,1965–1978.

Heino J, Tolkkinen M, Pirttilä AM, Aisala H, Mykrä H. (2014). Microbial diversity and community– environment relationships in boreal streams. Journal of Biogeography 41: 2234–2244.

Hugoni, M., Taib, N., & Debroas, D. (2013). Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proceedings of the National Academy of Sciences, 110(15), 6004-6009.

Jones, S. E., & Lennon, J. T. (2010). Dormancy contributes to the maintenance of microbial diversity. Proceedings of the National Academy of Sciences, 107(13), 5881–5886.

Jones, S. E., Cadkin, T. A., Newton, R. J., & McMahon, K. D. (2012). Spatial and temporal scales of aquatic bacterial beta diversity. Frontiers in Microbiology, 3, 1–10.

Kirchman, D. L. (2016). Growth Rates of Microbes in the Oceans. Annual Review of Marine Science, 8(1), 285–309.

Langenheder, S., & kely, A. J. S. E. (2011). Species sorting and neutral processes are both important during the initial assembly of bacterial communities. The ISME Journal, 5(7), 1086–1094.

Lapierre JF, del Giorgio PA (2014) Partial coupling and differential regulation of biologically and photochemically labile dissolved organic carbon across boreal aquatic networks. Biogeosciences 11: 5969–5985.

Lapierre JF, Seekell DA, del Giorgio PA. (2015). Climate and landscape influence on indicators of lake carbon cycling through spatial patterns in dissolved organic carbon. Globlal Change Biology; e-pub ahead of print 14 June 2015; doi:10.1111/gcb.13031.

Lear, G., Bellamy, J., Case, B. S., Lee, J. E., & Buckley, H. L. (2014). Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds,

1–12.

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., et al. (2004). The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters, 7(7), 601–613.

Lennon, J. T., & Jones, S. E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. Nature Reviews Microbiology, 9(2), 119–130.

Lennon, J. T., Aanderud, Z. T., Lehmkuhl, B. K., & Schoolmaster, D. R. (2012). Mapping the niche space of soil microorganisms using taxonomy and traits. Ecology, 93(8), 1867–1879.

Lindh, M. V., Sjöstedt, J., Andersson, A. F., Baltar, F., Hugerth, L. W., Lundin, D., et al. (2015). Disentangling seasonal bacterioplankton population dynamics by high-frequency sampling. Environmental Microbiology, 17(7), 2459–2476.

Lindström, E. S., & Bergström, A.-K. (2004). Influence of Inlet Bacteria on Bacterioplankton Assemblage Composition in Lakes of Different Hydraulic Retention Time. Limnology and Oceanography, 49(1), 125–136.

Lindström, E. S., & Langenheder, S. (2012). Local and regional factors influencing bacterial community assembly. Environmental Microbiology Reports, 4(1), 1–9.

Lindström, E. S., Forslund, M., Algesten, G., & Bergström, A.-K. (2006). External Control of Bacterial Community Structure in Lakes. Limnology and Oceanography, 51(1), 339–342.

Livermore, J.A., Emrich, S.J., Tan, J. & Jones, S.E. (2014). Freshwater bacterial lifestyles inferred from comparative genomics. Environmental Microbiology, 16, 746–758.

Liu, L., Yang, J., Yu, Z., & Wilkinson, D. M. (2015). The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. The ISME Journal, 9(9), 2068–2077.

Locey, K.J. & Lennon, J.T. (2016). Scaling laws predict global microbial diversity. Proceedings of the National Academy of Sciences, 113, 5970–5975.

Lomolino, M. V., B. R. Riddle, R. J. Whittaker, and J. H. Brown. 2010. Biogeography. Sinauer Associates.

Logares, R., Lindström, E. S., Langenheder, S., Logue, J. B., Paterson, H., Laybourn-Parry, J., et al. (2013). Biogeography of bacterial communities exposed to progressive long-term environmental change. The ISME Journal, 7(5), 937–948.

Logares, R., Mangot, J.-F., & Massana, R. (2015). Rarity in aquatic microbes: placing protists on the map. Research in Microbiologoy, 166(10), 831–841.

Logue, J. B., & Lindström, E. S. (2010). Species sorting affects bacterioplankton community composition as determined by 16S rDNA and 16S rRNA fingerprints. The ISME Journal, 4(6), 729–738.

Lozupone, C. A., & Knight, R. (2007). Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences of the United States of America, 104(27), 11436–11440.

Lutz, C., Erken, M., Noorian, P., Sun, S., & McDougald, D. (2013). Environmental reservoirs and mechanisms of persistence of Vibrio cholerae. Frontiers in Microbiology, 4, 375.

Lynch, M. D. J., & Neufeld, J. D. (2015). Ecology and exploration of therare biosphere. Nature Reviews Microbiology, 13(4), 217–229.

Maechler, M. (2015). diptest: Hartigan's Dip Test Statistic for Unimodality - Corrected. R package version 0.75-7. Available at: http://CRAN.Rproject.org/package=diptest. Last accessed 2016-05-01.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics (Oxford, England), 27(21), 2957–2963.

Magurran, A.E. (2007). Species abundance distribution over time. Ecology Letters, 10, 347-354.

Magurran, A.E. and Henderson, P.A. (2003). Explaining the excess of rare species in natural species abundance distributions. Nature 422, 714–716.

McCallister, S. L., & Giorgio, P. A. (2008). Direct measurement of the d13C signature of carbon respired by bacteria in lakes: Linkages to potential carbon sources, ecosystem baseline metabolism, and, and CO2 fluxes. Limnology and Oceanography, 53(4),104-1216.

McGeoch, M.A. and Gaston, K.J. (2002). Occupancy frequency distributions: patterns, artefacts and mechanisms. Biological reviews of the Cambridge Philosophical Society, 77, 311–331.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. PeerJ, 2(1), e593.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ, 3, e1420.

Matthews, T.J. & Whittaker, R.J. (2015). REVIEW: on the species abundance distribution in applied ecology and biodiversity management. Journal of Applied Ecology, 52, 443–454.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K. et al. (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology Letters, 10, 995–1015.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology, 4(2), 102–112.

Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D., & Horner-Devine, M. C. (2011). Drivers of bacterial beta-diversity depend on spatial scale. Proceedings of the National Academy of Sciences of the United States of America, 108(19), 7850–7854.

Mou, X., Sun, S., Edwards, R. A., Hodson, R. E., & Moran, M. A. (2008). Bacterial carbon processing by generalist species in the coastal ocean. Nature, 451(7179), 708–711.

Nelson, C. E., Sadro, S., & Melack, J. M. (2009). Contrasting the influences of stream inputs and landscape position on bacterioplankton community structure and dissolved organic matter composition in high-elevation lake chains. Limnology and Oceanography, 54(4), 1292–1305.

Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S. K., et al. (2011). Global patterns in the biogeography of bacterial taxa. Environmental Microbiology, 13(1), 135–144.

Nemergut, D. R., Shade, A., & Violle, C. (2014). When, where and how does microbial community composition matter? Frontiers in Microbiology, 5(497), 1-3.

Neuenschwander, S.M., Pernthaler, J., Posch, T. & Salcher, M.M. (2015). Seasonal growth potential of rare lake water bacteria suggest their disproportional contribution to carbon fluxes. Environmental Microbiology, 17, 781–795.

Newman, D. K., & Banfield, J. F. (2002). Geomicrobiology: how molecular-scale interactions underpin biogeochemical systems. Science (New York, NY), 296(5570), 1071–1077

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews : MMBR, 75(1), 14–49.

Niño-García, J. P., Ruiz-González, C., & del Giorgio, P. A. (2016). Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. The ISME Journal, 10(7),1755-1766.

Niño-García, J. P., Ruiz-González, C., & del Giorgio, P. A. The large-scale spatial distribution of bacteria across boreal networks differentiates the adaptive core versus accidental components of lake bacterioplankton. Ecology Letters, 19, 1506–1515.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MH, and Wagner H. (2015). vegan: Community Ecology Package. R package version 2.3-0. http://CRAN.R-project.org/package=vegan.

Pedrós-Alió, C. (2006). Marine microbial diversity: can it be determined? Trends in Microbiology, 14(6), 257–263.

Pedrós-Alió, C. (2012). The rare bacterial biosphere. Annual Review of Marine Science, 4, 449–466.

Pernthaler, J., & Amann, R. (2005). Fate of Heterotrophic Microbes in Pelagic Habitats: Focus on Populations. Microbiology and Molecular Biology Reviews, 69(3), 440–461.

Pommier, T., Canbäck, B., Riemann, L., Boström, K. H., Simu, K., Lundberg, P., et al. (2006). Global patterns of diversity and community structure in marine bacterioplankton. Molecular Ecology, 16(4), 867–880.

Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., et al. (2007). The role of ecological theory in microbial ecology. Nature Reviews Microbiology, 5(5), 384–392.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research, 41(D1), D590–D596.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ramette, A., & Tiedje, J. M. (2007). Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. Proceedings of the National Academy of Sciences of the United States of America, 104(8),2761-2766.

Ranjard, L., Dequiedt, S., eacute, N. C. P. E. V.-B., Thioulouse, J., Saby, N. P. A., Lelievre, M., et al. (1AD). Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. Nature Communications, 4, 1434–10.

Rasilo T, Prairie YT and del Giorgio PA. (2015). Large-scale patterns in summer diffusive CH_4 fluxes across boreal lakes, and contribution to diffusive C emissions. Global Change Biology, 21: 1124–1139.

Read, D. S., Gweon, H. S., Bowes, M. J., Newbold, L. K., Field, D., Bailey, M. J., & Griffiths, R. I. (2014). Catchment-scale biogeography of riverine bacterioplankton. The ISME Journal, 9(2), 516–526.

Ren, L., Jeppesen, E., He, D., Wang, J., Liboriussen, L., Xing, P., & Wu, Q. L. (2015). pH Influences the Importance of Niche-Related and Neutral Processes in Lacustrine Bacterioplankton Assembly. Applied and Environmental Microbiology, 81(9), 3104–3114.

Ruiz-González C, Niño-García JP, Lapierre JF, del Giorgio PA. 2015a. The quality of organic matter shapes the functional biogeography of bacterioplankton across boreal freshwater ecosystems. *Global Ecology and Biogeography* ; DOI: 10.1111/geb.12356

Ruiz-González, C., Niño-García, J. P., & del Giorgio, P. A. (2015). Terrestrial origin of bacterial communities in complex boreal freshwater networks. Ecology Letters, 18(11), 1198–1206.

Salcher, M. M. (2013). Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. Journal of Limnology, 73(1s), 74-87.

Saunders, A. M., Albertsen, M., Vollertsen, J., & Nielsen, P. H. (2015). The activated sludge ecosystem contains a core community of abundant organisms, 10(1), 11–20.

Savio, D., Sinclair, L., Ijaz, U. Z., Blaschke, A. P., Reischer, G. H., Bloeschl, G., et al. (2015). Bacterial diversity along a 2 600 km river continuum. Environmental Microbiology, 17(12), 4994-5007.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al.* (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* 75:7537–7541.

Shade, A., & Gilbert, J. A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. Trends in Microbiology, 23(6), 335–340.

Shade, A., Caporaso, J. G., Handelsman, J., Knight, R., & Fierer, N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time, 7(8), 1493–1506.

Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N., & Gilbert, J. A. (2014). Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity. mBio, 5(4), 1-9.

Shade, A., Peter, H., Allison, S. D., Baho, D. L., & Berga, M. (2012). Fundamentals of microbial community resistance and resilience. Frontiers in Microbiology, 3(417) 1-19.

Sjostedt, J., Koch-Schmidt, P., Pontarp, M., CANBÄCK, B., TUNLID, A., LUNDBERG, P., et al. (2012). Recruitment of Members from the Rare Biosphere of Marine Bacterioplankton Communities after an Environmental Disturbance. Applied and Environmental Microbiology, 78(5), 1361–1369.

Sloan, W. T., Woodcock, S., Lunn, M., Head, I. M., & Curtis, T. P. (2006). Modeling

Taxa-Abundance Distributions in Microbial Communities using Environmental Sequence Data. Microbial Ecology, 53(3), 443–455.

Souffreau C, Van der Gucht K, Van Gremberghe I, Kosten S, Lacerot G, Meirelles Lobão L, *et al.* (2015). Environmental rather than spatial factors structure bacterioplankton communities in shallow lakes along a >6000 km latitudinal gradient in South America. *Environmental Microbiology* 17, 2336–2351.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences of the United States of America, 103(32), 12115–12120.

Staley, C., Gould, T. J., Wang, P., & Phillips, J. (2014). Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. Frontiers in Microbiology, 5(414),1-11.

Steele, J. A., Countway, P. D., Xia, L., Vigil, P. D., Beman, J. M., Kim, D. Y., et al. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. The ISME Journal, 5(9), 1414–1425.

Székely, A. J., & Langenheder, S. (2014). The importance of species sorting differs between habitat generalists and specialists in bacterial communities. FEMS Microbiology Ecology, 87(1), 102–112.

Székely, A. J., Berga, M. E., & Langenheder, S. (2012). Mechanisms determining the fate of dispersed bacterial communities in new environments, 7(1), 61–71.

Torsvik, V., Øvreås, L., & Thingstad, T. F. (2002). Prokaryotic diversity--magnitude, dynamics, and controlling factors. Science, 226(5570), 1064-1066.

Ulrich, W. & Ollik, M. (2004). Frequent and occasional species and the shape of relative-abundance distributions. Diversity and Distributions, 10, 263–269.

Vachon, D., & del Giorgio, P. A. (2014). Whole-Lake CO2 Dynamics in Response to Storm Events in Two Morphologically Different Lakes. Ecosystems, 17(8), 1338–1353.

Van der Gucht, K., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., et al. (2007). The Power of Species Sorting: Local Factors Drive Bacterial Community

Composition over a Wide Range of Spatial Scales. Proceedings of the National Academy of Sciences of the United States of America, 104(51), 20404–20409.

Van der Heijden, M. G. A., Bardgett, R. D., & van Straalen, N. M. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology Letters, 11(3), 296–310.

Vergin, K. L., Done, B., Carlson, C. A., & Giovannoni, S. J. (2013). Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. Aquatic Microbial Ecology, 71(1), 1–13.

Villaescusa, J. A., Casamayor, E. O., Rochera, C., Velázquez, D., Chicote, A., Quesada, A., & Camacho, A. (2010). A close link between bacterial community composition and environmental heterogeneity in maritime Antarctic lakes. International Microbiology : the Official Journal of the Spanish Society for Microbiology, 13(2), 67–77.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology, 73(16), 5261–5267.

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: The unseen majority. Proceedings of the National Academy of Sciences of the United States of America, 95 (12), 6578–6583.

Wilhelm, L., Besemer, K., Fragner, L., Peter, H., Weckwerth, W., & Battin, T. J. (2015). Altitudinal patterns of diversity and functional traits of metabolically active microorganisms in stream biofilms, 1–11.

Williams, R. J., Howe, A., & Hofmockel, K. S. (2014). Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. Frontiers in Microbiology, 5, 358.

Yannarell, A. C., & Triplett, E. W. (2005). Geographic and environmental sources of variation in lake bacterial community composition. Applied and Environmental Microbiology, 71(1), 227–239.

Zillio, T. & Condit, R. (2007). The impact of neutrality, niche differentiation and species input on diversity and abundance distributions. Oikos, 116, 931–940.